Algorithmic Interventions:
The Power and Politics of Algorithmic Decision Systems

A DISSETATION SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA

BY

Jennifer A. Halen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

Co-Advisors: Dr. Benjamin Bagozzi and Dr. Joanne Miller

August 2019

Acknowledgments

I would like to thank my supportive committee members for their comments, support, and patience during what has been a tumultuous period both intellectually and personally.

First, I owe a tremendous debt of gratitude to Dr. Benjamin Bagozzi. Ben had the amazing ability to unfailingly provide both uplifting encouragement and rigorous, constructive advice. He was also a constant and accessible pillar of support even as both of our institutions changed. I could not have navigated this process without his consistent guidance and advice.

I would also like to thank Dr. Joanne Miller for her willingness to provide feedback and mentorship whenever I needed them. I am a different person from the one she called in Reno, NV to welcome to the graduate program, but her consistent support has been invaluable.

I am grateful for Dr. Raymond Duvall's longtime assistance, advice, and mentorship. I will forever be appreciative for his willingness to follow my intellectual curiosity, no matter where it led. I also appreciate his enthusiasm and support for my pursuits outside of traditional academic research.

I am also thankful for the opportunities that Dr. Claudia Neuhauser gave me. These practical and engaging experiences changed my intellectual and personal relationship with my subject matter, and I will be forever grateful. I can only hope to be engaged in similarly community-embedded projects throughout my career.

I would also like to thank member of the faculty for their helpful feedback and advice, in particular, Dr. James Hollyer, Dr. Timothy Johnson, and Dr. Kathryn Pearson. Similarly, I am grateful for my peers at UMN who were steadfast friends and colleagues throughout this experience. In particular, I will forever be grateful for Dr. Ore Koren, Dr. Christina Farhart, Dr. Kashif Riaz, Katrina Heimark, Jay Vargas, and Dalia Selman.

I also want to thank several professors at my undergraduate institution, the University of Nevada, Reno.

First, without Dr. Stacy Gordon-Fisher's invaluable mentorship, friendship, and willingness to include two eager undergraduates in her book project, I doubt I would be here today.

Dr. Robert Ostergard also offered significant support, feedback, and honest advice. I thank him for being a longtime mentor, colleague, and friend.

Finally, Dr. Thomas Smith sparked my intellectual interest in a way that became indistinguishable from who I am, and who I want to be, both as a person and as a scholar. His support and respect have meant the world.

I am also indebted to The UNR Ronald E. McNair Post-Baccalaureate Achievement Program which, as a first-generation college graduate, gave me the opportunities, support, and resources to be a successful graduate student. This dissertation would not be possible without this valuable program.

While at the University of Minnesota, my research was funded by the Clara Ueland Graduate Fellowship and The National Science Foundation Graduate Research Fellowship Program. The final two years of my research were enabled by a Fellowship at the Berkman Klein Center for Internet and Society at Harvard University.

While at the Berkman Klein Center I also gained invaluable friendships which challenged me personally and intellectually. I want to thank the 2017-2018 and 2018-2019 BKC cohorts. In particular, I would like to thank Dr. David Weinberger, Dr. James Wahutu, Aida Joaquin Acosta, Doaa Abu-Elyounes, Dr. Jonas Kaiser, Kathy Pham, Levin Kim, Nikki Bourassa, Jenna Sherman, Adam Nagy, Prof. Christopher Bavitz, Elena Goldstein, Juan Ortiz Freuler, Dr. Gretchen Greene, Elettra Bietti, Salome Viljoen, Jenny Korn, Dr. Jasmine McNealy, Dr. Nina Springer, Suchana Seth, Victoria Borneman, Yasodara Córdova, Dr. Padmashree Gehl Sampath, Dr. Jie Qi, Ram Shankar Siva Kumar, Dennis Redeker, Mariel García-Montes, Nikolas Guggenberger, Dr. Luke Stark, Ben Green, and Prof. Charles Nesson.

I would also like to thank my little sister, Jolene Bowers, for her support and patience throughout this very long process. Her humor, understanding, and enthusiasm have helped get me through a very long experience.

I also appreciate my Dad, Lynn Bowers, who passed away almost exactly one year before this dissertation was complete. Despite only having a partial high school education, he was the most avid reader I have ever known. Given that he hardly used technology beyond the radio, I have no idea what he would have thought of my research. I wish I had the chance to ask. Regardless, I will always be grateful for his support as I found my way as a student and scholar.

I will also always be grateful for the love and joy of my 4-legged family members, Henry and Widget. These two rescue pups sacrificed numerous walks and playtimes for this dissertation but still love me anyway.

Finally, and most importantly, I would like to thank my spouse and partner, Ryan Halen. He has made learning fun since we were in college. With him, academic work never felt like work, whether we were students together or scholars, and I will never be able to repay him for that. I have had the good fortune to have him by my side throughout our academic experiences and will always be grateful that he has been my constant

companion, peer, colleague, periodic tutor, frequent proofreader, gentle skeptic, and unwavering supporter. He has always made me a better person and a better scholar. For these reasons and so many others, I dedicate this dissertation to him. We have come a long way, and I can't wait to see what we do next.

Dedication

For Ryan. Finally.

Dissertation Abstract

Algorithmic Decisions Systems (ADS) are now commonly integrated within governing institutions ranging from the criminal justice system to social welfare programs. In an apolitical world, we might expect these considerations to be purely decided on bureaucratic optimization, but within highly politicized and contentious policy areas, we can expect each of these decisions to be opportunities for strategic interactions between interested parties. My dissertation seeks to address three core questions regarding governmental adoption of Algorithmic Decision Systems. 1) What attributes of these systems may lead legislators to support their use? 2) Does the inclusion of ADSs increase legislative support for bills that include these systems? 3) Do changes in the narrative around ADSs, particularly perceived public backlashes, impact legislative support? Throughout Chapters 1 and 2, I trace the history of ADSs as a natural evolution of bureaucratic systems and unpack the characteristics of ADSs that may be attractive to policymakers. In Chapter 3, I use a game-theoretic model to explore the way that ADSs are used to expand legislative control over bureaucratic decision-making. In Chapter 4, I use an empirical model to analyze legislative support for criminal justice legislation in U.S. state legislatures over the period 2012-2018, I provide evidence that suggests that inclusion of ADSs did increase some forms of legislative support for bills that included them, but that these effects were eroded and then overcome in later years by the recent critical turn and public backlash against ADSs. Lastly, I conclude my dissertation by discussing possible research avenues for future scholarly work on governmental adoption of Algorithmic Decisions Systems.

Table of Contents

List of Tables

List of Figures

List of Equations

Chapter 1
Algorithmic Interventions

Algorithmic Decision Systems (ADS) have grown into a new and expanding avenue to influence social policy. ADS are automated, data-driven processes used for decision-making purposes. These technical decision systems are automated processes that assess information and return an analysis. Generally, ADSs fall into two categories: "decision making systems" (i.e., fully automated ADS) and "Decision Aiding Systems" (i.e., systems which inform or guide)[1] (Selinger and Seager 2012). Algorithmic decision-making systems process information and then automatically issue a decision, such as credit scoring systems that automatically result in loan decisions. While, algorithmic decision-aiding systems produce information or "suggestions" that are intended to guide the human responsible for making the decision. For instance, while mobile health apps that track and suggest food and exercise do not explicitly require the user to follow a "healthy" routine, the apps recommend and endorse specific behaviors with the intention of altering the user's actions (Thaler and Sunstein 2009; Yeung 2017). ADSs are often software-based systems; however, the level of sophistication of ADSs range in complexity and technological sophistication. Technical approaches include (but are not limited to): simple rule-based systems, inferential statistical methods, machine learning

[1] Most examples discussed will focus on algorithmic decision aiding systems of varying strengths. Where appropriate, I will also use names of types of ADS systems, most often "risk assessments." Also, to aid in ease of approachability for scholars from different backgrounds, I also use the terms algorithmic policies and algorithmic decision frameworks interchangeably with ADS. For full definitions and descriptions of each term, see Chapter 2.

(ML) methods, and even applications of Artificial Intelligence (AI). In some circumstances, this term may also be appropriate for actuarial processes that do not require software. ADSs are ubiquitous in the private sector (e.g., Google's search results, Netflix's movie recommendations, Facebook Friend Suggestions, etc.) and have also been gaining momentum in public and governing spaces, ranging from courtrooms[2] to classrooms[3] and welfare offices[4].

Throughout the late 20th century and the very beginning of the 21st century, this software was viewed as a beneficial and objective approach to government decision making, emphasizing a turn toward "data-driven decisions" and "evidence-based policy" (Perri 2002, Marston and Watts 2003). However, as Algorithmic Decision Systems began to affect nearly every facet of life in much of the world, including what we read[5], what we watch,[6] where we work,[7] who we have friendships and relationships with,[8] and how we vote[9] to name only a few, the popular narrative began to shift rapidly. By the mid-

[2] Examples include algorithmic needs assessments to determine possible treatments or interventions for a defendant (such as substance abuse rehabilitation), algorithmic pretrial risk assessments that assess an arrestee's likelihood of missing a court date or committing a new crime if released on bail, and software used to aid decisions related to sentencing (Bechtel et al. 2017; Kehl, Guo, and Kessler 2017).
[3] Examples include algorithms that determine a child's assigned school (McKinley 2010), algorithms that aid in college admissions (Baig 2018), and risk assessments that alert administrators to students who may be at risk for dropping out (Berens et al. 2018).
[4] Examples include automated eligibility software that determines if an applicant will receive public assistance, such as Medicaid, food stamps, and cash-assistance, as well as the amount or level of assistance provided (Eubanks 2018).
[5] Examples include news aggregator programs like Apple's News app.
[6] Most content streaming services integrate some type of recommender system, including Netflix, Hulu, and Amazon Prime Video.
[7] For instance, many companies use ADSs to determine qualified candidates within a set of applicants (Raghavan 2019).
[8] Examples include Facebook or LinkedIn suggestions for new contacts and any number of dating websites and apps such as Match.com or Tinder (Eslami et al. 2014, Gillespie 2014)
[9] see Noble 2018 and Gillespie 2018 for a variety of examples of algorithmic impacts on ideological and political behavior.

2010s, the adoption of decision systems within national, state, and local government agencies sparked concerns regarding the ethical and legal implications of the design and implementation of the systems (see Mittelstadt et al. 2016). Headlines shifted from enthusiastic techno-optimism in the early 2010s[10] to dystopian descriptions of technology[11] by 2016.

The critical turn in the narrative about ADSs is indicative of the fact that they are, contrary to an 'objective' framing, inherently political artifacts. This shift has fueled a significant amount of scholarly interest in understanding the potential and observed impacts of deploying ADSs in social environments. This work has uncovered what should have been initially obvious: that ADSs are engaged in making political decisions based on, often unstated, values codified into the structure of these systems. ADSs are employed in a variety of situations that can involve determinations of who receives state resources and services, like welfare or other social aid, to who is likely to be at an increased risk of policing or incarceration; decisions which are inherently political in nature and cannot be determined without some pre-existing value system in place. The major goal of this dissertation is to elucidate some of the specific ways that ADSs manifest as political phenomena, foremost among them: how ADSs are a natural extension of the evolution of bureaucratic politics; how the public, media, and scholars engage with ADSs as an extension of debates over fundamentally political notions like

[10] For instance, Fast Company's 2012 headline "Fighting Violent Gang Crime With Math" (Coren 2012) which describes a predictive policing algorithm used by the LAPD.
[11] For instance, The New York Times 2017 article "Sent to Prison by a Software Program's Secret Algorithm" or The Economist's 2018 piece entitled "How data-driven policing threatens human freedom" (The Economist 2018).

fairness, ethics, and justice; and how ADSs can have potentially important impacts on explicitly political processes, like support for and passage of legislation.

Project Purpose and Motivation

Society is at a major inflection point in which citizens, legislators, and bureaucrats must decide how to respond to the real-world effects of digitization. Algorithmic processes are deeply ingrained in almost every facet of individuals' lives in much of the world, and many people and organizations are becoming more aware of the significance of computational technologies in the physical world (Crawford 2017; Mayson 2019). In particular, government adoption of ADSs has caused significant concern (Werth 2019; Mayson 2019). In addition to the broad public implications of structural and procedural changes to any government institution, the applications of ADSs have tended to have a greater effect on marginalized and vulnerable populations (Eubanks 2018). For example, in just a few of the government cases noted previously (see footnotes 3-5), algorithmic decision systems have become ingrained in school assignment decisions meant to combat racial segregation (McKinley 2010), judicial processes rife with economic and racial prejudice during pretrial detainment decisions (Bechtel et al. 2017; Kehl, Guo, and Kessler 2017), and public service processes for decisions regarding requests for basic nutrition assistance and advanced medical needs (Eubanks 2018).

Dreams of automated technologies bringing in a new age of justice and equality have been overshadowed by concerns for computational technology's harmful shortcomings. An influential exposé was published in May 2016 by ProPublica, an

investigative journalism organization. This exposé argued that a popular sentencing algorithm called COMPAS.[12] was "biased against blacks." Specifically, the authors found that COMPAS was twice as likely to assign high risk scores to black defendants in comparison to their white counterparts despite, on average, black defendants being significantly less likely to commit future crimes (Angwin et al. 2016). While the ProPublica article received significant pushback from the creator of COMPAS, Northpointe Inc. and others (see Corbett-Davies et al. 2017), it also spurred academic and industry interest in the topic of fairness in algorithmic systems (Courtland 2018). This critical narrative quickly spread to the popular press (Crawford 2017, Whittaker et al. 2018).

Algorithmic decision systems can initially seem like a niche concern for so much attention; normally this level of scrutiny is not applied to a specific policy tool, but rather larger policy areas with multiple competing policy prescriptions. However, ADSs are unique in that they are a relatively new policy intervention (in their current form) that has nonetheless found rapid success at being introduced and enacted in various legislative areas. For example, and I will discuss this at length in Chapter 4, over the preceding 7 years, in the field of criminal justice policy alone, 59 pieces of legislation involving ADSs were successfully passed and enacted across 46 states in the U.S; an additional 57

[12] COMPAS is an algorithm that analyzes a number of demographic, social, and physiological factors in a publicly unknown way (its underlying processes and weights are proprietary information) to derive an estimate for how much of a 'risk' an arrestee is for fleeing from their court date or committing a new crime if released on bail. This risk score is presented to a judge at a pretrial bond hearing (discussed further in Chapter 4) with the intention of informing the judge's decision as to whether to release an arrestee and under what conditions (see Northpointe 2015).

were introduced in 2018 alone, with 9 of them eventually being enacted. These systems are being deployed in situations that intimately affect people's daily lives, from determining child protective service investigations to who gets released from jail. These systems are potentially affecting millions of individual's lives in ways that represent some of the most common interactions between citizens and their respective states.

However, as important as these interactions and micro-level outcomes are, the implementation of technologies like ADSs can irrevocably change a policy landscape, ultimately making it more difficult to make future progress (see Whittaker et al. 2018). On the other hand, implementation of ADSs may break up long-standing, path-dependent processes, opening up the opportunity for changes that otherwise would not be able to occur. Furthermore, as I will discuss in Chapter 2, ADSs may serve to improve policy outcomes relative to currently biased or discriminatory status quos (Mayson 2019). However, scholars and policy observers also contend ADSs may only serve to obfuscate and reify existing prejudices and harms. The state of the literature on algorithmic policies is rife with these types of competing claims, and the stakes are high. Further cementing harmful systems with the veneer of scientific and technological objectivism would be a major normative harm. But, if this is not the case, is the repeated assertion that any implementation is harmful having a dampening effect on our willingness to pursue what is a viable tool for social progress?

Destructive trends can appear on both ends of the spectrum. In one scenario, the negative narrative could produce a chilling effect on the adoption of ADSs in government or even on the development and innovation within machine learning and AI research.

This scenario already occurred more than once. Due to unmet expectations, funding, and interest in the subject of artificial intelligence slowed leading to a series of "AI Winters" in the 1970s, 80s, and 90s (Elish and boyd[13] 2018). In a differing, but equally concerning scenario, the public could become exhausted with competing narratives and lose interest in creating more ethical expectations and regulations for the use of an ever-growing technology (Crawford 2017, Whittaker et al. 2018). It is exactly these types of competing claims that require a robust, empirical literature to disentangle competing claims and properly identify the relative sizes of various relationships, impacts, and trends. Yet, the topic of algorithms and public policy is still very much nascent, which makes this stage of research critical. Research being conducted now will set the tenor for how future discussion of ADSs are handled by researchers, so it is imperative to pair important conceptual and theoretical work with empirics when possible to strengthen the evidence and literature on the nature and impacts of ADSs. This is a second major goal of this dissertation: to emphasize the need for sociotechnical issues, such as algorithmic decision systems, to be situated within prior understandings of political institutions and processes.

State of the Field and Research Objectives

There is a growing research literature on the social and ethical implications of public Algorithmic Decision Systems. Researchers have examined the potential for inaccurate and discriminatory predictions (Barocas and Selbst 2016, Zarsky 2016; Eubanks 2018), the legality of the use of ADSs in public and private contexts (Doshi-

[13] stylized lowercase

Velez and Kortz 2017), issues of transparency and privacy (O'Reilly 2013, Pasquale 2015, Annany and Crawford 2018), and mechanisms for accountability if an algorithmic system results in harm (Diakopoulos 2013; Buhmann, Paßmann, and Fieseler 2019). This literature is characterized by two significant trends. One, it is heavily conceptual in nature; what empirical evidence exists tends to be qualitative information gained from interviews with individuals from targeted populations, or quantitative information obtained via surveying targeted populations (See Werth 2019). This is important work, both the purely conceptually and empirically, but it is overwhelmingly focused on micro-level interactions, and fails to capture meso and macro-level[14] evidence and relations. For a subject that is concerned with systemic level biases and impacts, this is a major space for scholarly attention.

Second, the literature is overwhelmingly comprised of analyses focused on the outcomes of ADSs, either in the form of understanding additional effects/externalities experienced by targeted populations (O'Neill 2016, Eubanks 2018, Noble 2018) or in the form of validation studies that seek to determine how accurate ADSs are with their predictions (Cohen and Lowenkamp 2019; see Desmarais, Johnson, and Singh 2016 for a meta-analysis). Understanding outcomes in undoubtedly important, but as inherently political artifacts, ADSs have potential implications for political processes, and understanding these impacts will be just as important as understanding outcomes when

[14] Meso and macro-level phenomena describe the fundamental location of an interaction of interest. Contrasted with micro-level, which focuses on the interactions of individuals, macro-level phenomena focus on large systems and structures, like a legislature of policy area, while meso level phenomena can describe either bridging interactions between individuals and larger structures or interactions within and between subsystems of larger macro-level systems.

determining the true impacts of ADSs, their likely future, and how societies characterized by the widespread adoption of algorithmic policies may change over time.

Due to ADS' scalability and ability to rapidly synthesize information, algorithmic decision systems can theoretically provide policymakers and other public officials with replicable, empirical predictions while also reducing human subjectivity and lowering the costs of often-overburdened governing agencies (Mayson 2019). Opponents[15] who oppose the use of decision systems in government largely focus on the explicit and implicit beliefs and motivations that become deeply ingrained in the design of technology, arguing that this can exacerbate current inequalities under the guise of objectivity and scientific rigor (O'Neil 2016). The frequent implementation of these systems in the lives of society's most vulnerable, such as defendants in the criminal justice system or at-risk children, have caused significant concern over the lack of oversight or regulation of this type of software (Kirchner 2015). These debates are important and have real effects on the creation and perception of algorithmic decision-making tools and the broader pieces of legislation they are attached to. However, the choice to use these decision systems is not made in a vacuum, but rather implemented within complex policy environments by actors with political and institutional motivations and constraints.

[15] Opponents to algorithmic interventions in any given policy area can include policymakers, advocates, organized interests, and citizens. Opposition can arise from either concerns about the ethics and legality of governmental ADSs, or entrenched interests threatened by the employment of ADS (such as the bail bond industry in the context of pretrial risk assessments, discussed at length in Chapter 4).

The choice to support or oppose the adoption of an algorithmic intervention[16] is not only based on a policymaker's independent policy goals but also the expected behavior of others in the issue area's policy making space. Therefore, in a traditional spatial framework, the support or opposition for algorithmic interventions is also based on whether or not this decision is believed to move the policy outcome closer or farther away from the legislator's ideal point, which is embedded in a complex, inherently political system. This belief will include a policymaker's independent policy goals, the expected behavior of others in the issue area's space, and the perceived alternatives to implementing an algorithmic intervention. This setup informs the rationale for the formal model developed in Chapter 3.

Continuing in this vein, I analyze how algorithmic interventions affect the policy process. This is done through chapters 3 and 4 in three specific ways. The first is to unpack some of the possible motivations legislators have for employing decision systems in public, governing institutions through the use of a game-theoretic delegation model. What is it that legislators believe they will achieve when mandating the use of ADSs in legislation? Next, I explore the dynamics that prompted the sudden turn from optimism about decision aids to pessimism and critique. These two facets inform two general hypotheses that form the basis of the final empirical analysis. From the theory discussion in Chapter 2 and the formal model analysis in Chapter 3, I make the argument that there is a fundamental structural advantage for ADSs, relative to other policy tools, when

[16] I use "Algorithmic Interventions" to refer to the choice to use an algorithmic tool, such as decision systems, as a part of a public policy solution.

included in legislation that may make algorithmic interventions appealing to legislators. However, the nature of the critical turn has generated a backlash against such systems that may have eroded these benefits. Tests regarding positive baseline support of algorithmic interventions and the potential effect of the subsequent backlash, form the basis of the empirical investigation in Chapter 4.

Primary Objectives and Themes

The discussion thus far has focused on the three primary objectives of my dissertation: 1) demonstrating the inherently strategic, political nature of algorithmic decision systems, 2) grounding the discussion over ADSs and society in both a conceptual and empirical manner, and 3) focusing on how ADSs have potential impacts on political process as well as outcomes. One of the major themes I will discuss throughout this dissertation focuses on the tension between systemic, structural advantages and the unstable nature of public scrutiny. Much of the public work involved in the critical turn and resulting backlash against ADSs has relied on attracting media and public attention on ADS usage and adoption by highlighting continued disparities in outcomes and making appeals to fairness and human discretion. However, these issues are highly context dependent in that they are dependent on the specific nature of the algorithmic interventions and broader nature of public attentiveness and attitudes around these issues. As algorithmic policies change and adapt, as the public engages and interacts with them more often, and as more issues continue to compete for the attention of the public, media, and various interest groups, the context around algorithmic

interventions is just as likely to change. However, systemic advantages, if they exist, are only likely to change if something fundamental about the system changes. Therefore, for scholars to gain a proper understanding of the continued, if any, role of ADSs and the impacts they may have on society, it is important to understand these potential systemic advantages, and by extension how ADSs may impact political processes.

Dissertation Outline

Chapter 2

In the next chapter I will unpack my conceptual framework and discuss key concepts. I begin by examining the historical roots of the quest for stable, impartial, rule-based governmental procedures (e.g., the concept of the "rule of law") and the trajectory leading to the modern bureaucratic state, the popularity of "evidence-based policy" and, eventually, algorithmic decision systems. In this way, I establish ADSs as a natural evolution of the same explicitly political processes and tensions that inform the nature of bureaucratic politics and the administrative state. Next, I will connect the characteristics (or the perceived characteristics) of algorithmic decision systems with the American political process in order to illustrate the motivations and goals that may have led to the rapid increase in the digitization of governmental decision processes. I then discuss the causes and consequences of the expanding oppositional narratives, which claim that inclusion of algorithmic decision systems will only reify and expand problems of inequality. I conclude by summarizing these concepts and discussing their relationships to the following chapters.

Chapter 3

Chapter 3 uses a game-theoretic model in order to understand how characteristics of decision systems, specifically the prospect of enabling a legislator's indirect political control of policy implementation, may motivate the legislator to support policies that include decision systems. Previous research has emphasized the role that bureaucratic behavior has in influencing policy making. This is because of a fundamental principal-agent problem[17] within policymaking. That is, the legislator (i.e., the principal) must allow some on-the-ground discretion within legislation in order to account for real-time changes or unexpected contexts. However, increased discretion given to bureaucrats and other practitioners comes with an increased risk that they will use this discretion in a way that is contrary to the legislators' goals or intentions (either for political or other reasons) (McCubbins et al. 1987, Epstein and O'Halloran 1999, Bendor and Meirowitz 2004, Gailmard 2009). Using a game-theoretic model informed by the canonical approach taken by Epstein and O'Halloran (1999), I examine the ways in which decision systems could be viewed as a tool to moderate bureaucratic discretion and increase legislative control over policy implementation. My analysis expands on previous formal models that examine the competing legislative goals of political control versus policy flexibility in the face of unpredictable change. The game-theoretic approach models the strategic interactions between the legislative principal and the bureaucratic agent. It expands on previous models by allowing the legislature to include an algorithmic constraint that

[17] That is, the agent (i.e., the bureaucracy) can act on behalf of the principal (i.e., the legislature) (Weingast 1984, Krause 2010).

changes over time (corresponding with a machine learning algorithm's ability to adapt given new information). This algorithmic constraint can be applied with varying authority (i.e., how binding the algorithmic output is on the policy). The findings suggest that under normal conditions[18], legislators may be incentivized to support legislation that includes algorithmic decision systems. This may be, in part, due to the characterization that the decision system will expand the legislators' political control over final on-the-ground policy outcomes by utilizing the algorithm to dynamically constrain bureaucratic discretion and keep realized policy outcomes closer to the legislator's preference. The findings from this model provide one of the major conceptual pillars for ADSs having structural advantages. To the extent that ADSs help solve, or at least improve, the principal-agent problem at the heart of delegatory decisions, they provide a clear advantage over other similar policy tools. These advantages can potentially help propagate ADSs in other legislative areas and maintain their usage over time.

Chapter 4

In Chapter 4, I explore how inclusion of an ADS relates to support for legislation. Specifically, I examine the impact of risk assessment algorithms on successfully passed criminal justice legislation across all fifty states for the period 2012-2017, and introduced

[18] Normal conditions in the context of a game-theoretic model refers to under conditions of equilibrium, i.e., when all actors in the game are playing their optimal strategies, and changes by any player would result in less preferable outcomes for that player. There is significant discussion, not previewed here, concerning how 'true' a constantly optimized equilibrium is in the real world, i.e., actors can make mistakes or incorrectly identify their preferences. I sidestep these very valid concerns by focusing on the situation where the actors recognize their goals and what strategies are most likely to achieve those goals, i.e., normal conditions.

criminal justice legislation, regardless of passage, in 2018. I do so by analyzing three outcome variables that reflect legislative support: bipartisan support, number of co-authors, and eventual enactment (only for 2018 data).

Pretrial risk assessments offer a useful test for both the systemic advantages as well as the temporal nature of public scrutiny. This is because pretrial risk assessments have seen a significant number of attempts at adoption throughout criminal justice legislation over the past decade, but also because these risk assessment algorithms serve as one of the main focal points for the critical turn and eventual backlash against ADSs. However, it is important to recognize that each algorithmic decision system is embedded within a unique context and these contexts also help establish the fundamental political nature of ADSs more broadly. Rather than analyzing a defendant's past and projecting the likelihood of violent crime in the future, an algorithm could, theoretically, assess different rehabilitation programs to better understand how to produce positive outcomes for defendants, either individually or as a population, in the future. However, each of these problem frames is only approachable through certain kinds of methods and models. Each design decision further refines and shapes what the model is fundamentally capable of, as well as what the outputs will be, but this level of technical sophistication is often obscured to both policymakers and the general public (Hannah-Moffat 2015, Elish and boyd 2018). This represents a gap between the ideation of an algorithmic decision system and the mechanisms that will be responsible for translating these intentions into the software that will be implemented. Through the use of pretrial risk assessments as a focal empirical test for the structural advantages and public scrutiny of ADSs, I am able to

jointly advance all three of the major goals of the dissertation: explicit political-ness, inclusion of empirical evidence, and focus on political processes.

Chapter 5

Chapter 5 concludes by summarizing the findings of the prior chapters, contextualizing the findings within the broader algorithm and society literature, and outlining future research avenues. The increasing adoption of decision-aiding technologies and the quick pace of technical advancement require social scientists to better understand technology as a part of the policy process and the sociopolitical consequences of that technology's implementation.

Prior to outlining a potential research agenda, I reflect on the major theme of the dissertation. I focus the discussion on how focusing research on explicitly understanding the politics and process implications of ADSs in an empirical manner can provide valuable insights into the major questions and concerns facing scholars, practitioners, and affected individuals in the field. Only by understanding these larger meso and macro-level facets of ADSs and politics can we build robust, transparent systems that leverage the potential benefits of such systems while providing mechanisms for identifying and addressing potential drawbacks.

Chapter 2
Theoretical Discussion and State of the Literature

This chapter will unpack my conceptual and theoretical arguments regarding potential systemic advantages of algorithmic interventions. I begin by examining the intellectual and institutional roots of the pursuit for a stable, impartial, rule-based government (i.e., the concept of the "Rule of Law") and the trajectory leading to the modern bureaucratic state, the popularity of "evidence-based policy" and, eventually, algorithmic decision systems. In this way, I establish ADSs as a natural evolution of the same explicitly political processes and tensions that inform the nature of bureaucratic politics and the administrative state. Next, I will connect the characteristics (or the perceived characteristics) of algorithmic decision systems with the American political process in order to illustrate the motivations and goals that may have led to the digitization of governmental decision processes. I then discuss the causes and consequences of the expanding oppositional narratives, which claim that inclusion of algorithmic decision systems will only reify and expand problems of inequality. I conclude with a discussion of key concepts and terms and the next steps within this dissertation. Connections between these ideas and my research questions, hypotheses, and methodology will be discussed throughout.

The Rule of Law and the Administrative State

The idea that government should consist of stable, impartial, rule-based procedures is fundamental to the western political tradition. Termed "the Rule of Law,"

this concept requires that all individuals and entities are "accountable to laws that are publicly promulgated [and] equally enforced" (Tolbert and Solomon 2006). This philosophy is discussed by numerous political theorists dating back to at least Aristotle and underpins much of foundational early American political thought. Political observers and intellectuals, such as John Locke (1690/2016) and Thomas Paine (1776/2003), emphasized the rule of law's contrasts to monarchy's divined superiority and arbitrary use of power (Tamanaha 2004, O'Donnell 2010).

Despite its long intellectual history, the execution of the ideals of stable, impartial, rule-based procedures have been difficult to perfect. In modern governments, these principles have been pursued through the creation of an extensive administrative state and bureaucratic system (Beetham 1996). Max Weber argued that bureaucratic administration "increased calculability" of the state, decreasing the likelihood of unpredictable interactions between the citizen and the state (Weber 1968; also see Vogl et al. 2019). Bureaucracies also lessened shocks of institutional and political changes by creating non-elected positions to administer laws and create processes to retain and share information (Redford 1958, Dunleavy et al. 2006, Vogl 2019). Over the latter part of the 20th century, the adoption of basic information technologies strengthened these characteristics by increasing storage capacity (elongating institutional memory) and expanding informational accessibility to both other state institutions and citizens (Landsbergen and Wolken 2001, Dunleavy et al. 2006).

The Development of Modern Bureaucracy

Bureaucratic routines and structures have lent some stability and consistency to state-individual interactions, they are imperfect mediums of the ideals of the Rule of Law. Some of this imperfection can be attributed to the fact that the institutions are rule-based, but the interpretation and implementation of those rules are reliant on the inconsistent use and misuse of judgment and discretion by human agents (Maynard-Moody and Portillo 2010). Despite consistent procedural mechanisms and routines, each human decision-maker is apt to respond in a different way to the same information, and the same decision-maker may even have difficulty evenly responding to similar situations with contextual differences (See Kahneman 2011, Bargh et al. 2012). This variation *within* and *between* individuals is due to complex and individualized cognitive processes and biases[19] that may result in implicit prejudices being integrated into policy implementation (Dovidio et al. 2002). Cognitive biases may be exacerbated by the fact that public officials often make dozens of life-altering decisions daily, which requires a fast pace that may increase the likelihood of errors or biases in decision-making going unnoticed. Bureaucrats may also either purposefully manipulate or modify procedures to align with their own preferences (known as bureaucratic drift) (Shepsle 1992; Béland, Rocco, and Waddan 2016) or to accommodate unreasonable demands (Lipsky 1980; Soss, Fording, and Schram 2011) or the slippage between a complex individualized

[19] Cognitive biases are unconscious, systematic errors in thinking processes ((Dovidio et al. 2002; Kahneman 2011, Bargh et al. 2012).

situation and a simplistic state categorization scheme (Prottas 1970, Maynard-Moody and Portillo 2010).

Discretion as Political Power

Early theories about bureaucratic power and discretion focused exclusively on high-ranking administrative elites, with no attention to frontline staff (Stein 1952), but subsequent research, starting with Michael Lipsky in the 1970s, found that "street-level bureaucrats[20]" may have significantly more power than most, or perhaps any, other actors in the system. Lipsky (1980) described these on-the-ground government employees as "the ultimate policymakers" because they are the arbiter of the last interaction and thus the final say of any citizen-state interaction. As Handler (1990) states

> "[d]espite the masses of legislation, rules, regulations, and administrative orders, most large, complex administrative systems are shot through with discretion from the top policy-makers down to the line staff—the inspectors, social workers, intake officers, police, teachers, health personnel, and even clerks. How they interpret the rules, how they listen to the explanations, how they help the citizen or remain indifferent, all affect the substance and quality of the encounter, an encounter made increasingly important because of our widespread dependence on the mod-ern state." (3–4)

Indeed, as Soss, Fording, and Schram (2011) argue, discretion "is ineradicable" in that bureaucrats "almost always know some way to push a decision in a preferred direction" (225). However, the authors note that the ways in which frontline bureaucrats[21]

[20] Street-Level Bureaucratic Theory (SLBT) refers to administrative discretion and judgment administered by government agents on the "front lines" (i.e., in direct contact with citizens), including police officers, social workers, teachers, and lower-level judges and court administrators (Lipsky 1989; also see Maynard-Moody and Portillo 2010).

[21] This study focused specifically on welfare caseworkers.

can exercise discretion varies and that the bureaucrats' "preferred direction" is often both responsive to policy demands and beneficial to the citizens with whom they are interacting (225-226).

Bureaucratic discretion is also embedded in the need to transform messy social situations into clean, generalizable categories defined in policies (Lipsky 1980, Prager 2007). For instance, in a welfare policy requiring unemployment beneficiaries to be "actively seeking employment," a caseworker could decide if rejecting certain employment given personal circumstances, such as costs or availability of childcare, violates this requirement (Soss, Fording, and Schram 2011). Similarly, many policies would be impractical to apply with limited time or resources. For instance, one police officer may not be capable of stopping every vehicle going even one mile over the speed limit, so they must decide which vehicles they will stop versus which they will ignore (Skolnick 1966; Brown 1981).

While some scholars and activists emphasize the need for, and benefit of, bureaucratic discretion and human empathy and judgment within state-citizen interactions, others have voiced staunch opposition to on-the-ground flexibility, arguing that it violates democratic hierarchies by subverting the will of elected officials or that it provides avenues for abuses of power and barriers to accountability (see Workman, Jones, and Jochim 2010). These opponents have sought to pinpoint and decrease, or even eradicate, openings for human fallibility and corruption within government systems.

Political Controls and Constraints

Scholars studying this topic from a legislature-centric perspective have largely viewed legislative delegation to the bureaucracy as a principal-agent relationship in which the agent (i.e., the bureaucracy) can act on behalf of the principal (i.e., the legislature) (Weingast 1984, Krause 2010). Problems arise when the agents act in a way contrary to the interests of the principal (McCubbins et al. 1987, Besley and McLaren 1993, Epstein and O'Halloran 1999, Bendor and Meirowitz 2004). This research emphasizes that deviation from the legislature's policy intent may be a consequence of politically or ideologically motivated bureaucrats who may intentionally sabotage or subvert the execution of the policy (Gailmard 2002), the ability for the agents to understand and interpret both the intention of the policy and how to best achieve its ends (Bendor and Meirowitz 2004), or to the institution's capacity which may not give the agency the needed resources to effectively implement the policy (Huber and McCarty 2004). Each of these circumstances requires different actions to successfully overcome the issue and achieve the policymaker's desired outcome.

Policymakers can attempt to confine bureaucratic agency by legislating rules and procedures that require, monitor, or enforce certain actions. However, even if policymakers could structure rigid enough mechanisms to control bureaucrats completely, there are many reasons why this would be harmful to policy outcomes. Legislators recognize that laws must be generalizable enough to be applied to many contextually different cases. However, legislators also want to preclude bureaucrats from inappropriately using this flexibility to depart from the intention of the original policy

(Bawn 1995). Thus, policymakers must engage in a strategic balancing act to manage the benefits and challenges of bureaucratic delegation (Krause 2010).

Unfortunately for legislators, balancing bureaucratic independence and policy control can be difficult to achieve. For instance, too much political control over a bureaucrat's actions limits their ability to gather and react to new information but giving bureaucrats too much agency may result in significant policy drift[22]. Furthermore, any increased uncertainty regarding bureaucratic competency may stymie policymakers' willingness to support legislation in a given area because of their uncertainty regarding ultimate policy outcomes (Noll 1983). Similarly, bureaucratic organizations with fewer resources (that may not have the staff, knowledge, or material resources to effectively execute a policy), can lower the incentives for policymakers to reform the policy area given the uncertainty about the bureaucrats' willingness or ability to comply with any set policy demands (Huber and McCarty 2004). And, as discussed above, extending control over, or even knowledge of, street-level bureaucrats' use of discretion includes further challenges.

Technology in Bureaucracy

The most straightforward mechanism for achieving bureaucratic constraints is for legislators to adopt increasingly rigid procedures and reporting requirements. However, some studies have shown that highly rigid structures are often impractical to implement

[22] Policy drift refers to policy change that occurs outside of formal legislative reform (Béland, Rocco, & Waddan 2016)

with finite resources and difficult to monitor given the vast number of state-citizen interactions (Prottas 1979, Brown 1981). This challenge has been met more recently by the adoption of surveillance technologies that can better monitor and track large numbers of cases. These technologies are often directed at both government agents, such as in the case of police dashcams or automated welfare case management systems (Brown 1981; Soss, Fording, and Schram 2011), and the public, through systems like employment monitoring and analysis of public service use (Gilliom 2001, Hannah-Moffat 2015, Eubanks 2018, also see Maynard-Moody and Portillo 2010).

Algorithmic Decision Systems as Bureaucratic Constraints

I argue that Algorithmic Decision Systems are a new kind of policy alternative that may be perceived by policymakers as a mechanism to constrain bureaucratic discretion while maintaining the intent of the original legislation and aligning with the ideals of the Rule of Law. In addition to the decision to fully automate a decision, ADSs can also be designed to influence or "guide" a decisionmaker using mechanisms that steer individuals toward making a certain decision, but does not explicitly constrain their behavior in any way (Thaler and Sunstein 2008). Similarly, as computational tools, they provide a record to monitor or evaluate bureaucratic actions, extending the scope of other information technology systems by not only recording decisions but also comparing it to a predicted or advised alternative (Hannah-Moffat 2015).

Additionally, ADSs may be perceived as providing other institutional benefits, as well. For instance, computational processes can synthesize and analyze vast quantities of

data to provide predictive outputs, thus increasing bureaucratic agency capacity by speeding up traditional information-seeking processes and more efficiently directing agency resources, such as limited time and staff (Mayson 2019, Vogl et al. 2019). These algorithmically produced results would also potentially enable bureaucrats to respond to contextual differences in predictable ways despite the increasing societal and institutional complexity of the modern world (Vogl et al. 2019).

I apply the legislative pursuit of controlling bureaucrats and the idealization of stability, impartiality, and rule-based systems to analyze my first research question:

> *Research Question 1: What Motivates Policymakers to Support Algorithmic Decision Systems?*

(Mis)Perceptions of Algorithmic Decision Systems

Algorithmic processes are generally not well understood by either policymakers or the general public (Domingos 2012; Pasquale 2015). In addition to the incentives described above, this gives ADSs an amount of vagueness or malleability when discussed as policy solutions that may alter traditional political coalitions of support. This is important with regard to my second research question:

> *Research Question 2: Do the Perceived Features of Algorithmic Decision Systems Increase Policymaker Support for ADS-Inclusive Legislation?*

Some of the misconceptions or lack of technical understanding about algorithmic processes likely stems from the pervasive misconceptions about the current attributes and capabilities of predictive software (O'Neil 2016, Elish and boyd 2018). For decades, tech companies have profited off of the notion that their technical solutions could fix

problems. Elish and boyd (2018) argue that this narrative went so far as to make computational and data-driven tools seem fantastical in a way that promotes the belief that they are "unknowable and inscrutable," which ultimately functions to relieve the company of having to validate their tools. Marketing materials go so far as to offer Machine Learning (ML) and Artificial Intelligence (AI) services that "work like magic" (Selbst, 2017; Elish and boyd 2018).

So, it follows, that if one believes that a tool is capable of producing objective, unbiased predictions (i.e., predictions that produce the correct or "right" outcome), they are likely to believe that said outcome will align with their own predictions, given that prior studies have shown individuals, and particularly "issue experts", as having a high level of confidence in their own prediction capabilities (Tetlock 2017, also see Kahneman 2011). In other words, policymakers may support legislation that includes algorithmic aids because they believe the policy will be pulled toward their own preferences. Opposing interests may believe this while simultaneously supporting the same language regarding "data-driven decisions." I argue that this may enable increased bipartisan support for bills that include Algorithmic Decision Systems. I test this theory in Chapter 4 by analyzing the relationship between inclusion an Algorithmic Decision Systems and bipartisan support, in the form of co-sponsorships, of legislation on Pretrial Reform nationwide.

The Critical Turn: Public Backlash and Popular Narratives

By the mid-2010s, the adoption of decision systems within national, state, and local government agencies had sparked concerns regarding the ethical and legal implications of the design and implementation of the systems (see Mittelstadt et al. 2016), particularly in contrast with the narrative of objectivity previously used. Much of the public work involved in the critical turn and resulting backlash against ADSs has relied on attracting media and public attention on ADS usage and adoption by highlighting continued disparities in outcomes and making appeals to fairness and human discretion. However, these issues are highly context dependent in that they rely on the specific nature of the algorithmic interventions and broader nature of public attentiveness and attitudes around these issues. As algorithmic policies change and adapt, as the public engages and interacts with them more often, and as more issues continue to compete for the attention of the public, media, and various interest groups, the context around algorithmic interventions is just as likely to change. This calls for an analysis that is a subset of the previous research question:

> *Research Question 2A: Do fluctuations in the narrative about Algorithmic Decision Systems alter levels of Legislative Support?*

Chapter 4 uses the substantial shift in academic and popular narratives regarding Pretrial Risk Assessment Algorithms, a type of ADS, to analyze the impact of variations in public discourse over a key time period on support for criminal justice bills that include such assessments.

Key Concepts and Terminology

Algorithms

An algorithm, in its most basic form, is a set of rules and procedures that refine or synthesize information for problem-solving purposes. Algorithms must have clearly specified rules and a finite number of steps to achieve a result (Chabert 1999, McKelvey 2014). For instance, simple mathematical methods, such as long division, are algorithms. These sets of rules do not need to be computational, but computers are faster and more reliable at executing explicit rules. When the rules become more complex and involve more information, computers process that information in a fraction of the time it would take a human. Furthermore, the rapid proliferation of algorithmic systems in the public sphere is only possible given computational advancements and a shift by individuals, businesses, and governments to turn to increasingly digitized processes (Buhmann, Paßmann, and Fieseler 2019).

Machine Learning and Artificial Intelligence

Machine learning refers to an advanced computational algorithm that can respond to diverse data semi-autonomously. The key feature of machine learning algorithms is that they independently adapt and iterate based on patterns detected within a dataset. There are three primary subtypes of learning algorithms: supervised learning, unsupervised learning, and reinforced learning. Put simply, supervised learning provides the algorithm with a large, pre-labeled dataset. As the algorithm goes through the data, it

detects similar attributes associated with each different data category. It can then apply these patterns to unlabeled data to predict its category. Unsupervised learning, however, is not given pre-labeled data. Instead, the algorithm analyzes unstructured data that has not been categorized in any way and is often used for exploratory analysis and dimension reduction. Reinforced learning methods use dynamic feedback (yes/no) after an action/decision is made to train the algorithm to behave in a certain way.

These systems, while not infallible, can rapidly analyze vast quantities of information and often provide useful insight into complicated patterns. They are often employed to predict the likelihood that a given case falls into a specific category. For instance, criminal risk assessments analyze a defendant's profile (usually containing data points regarding criminal history, suspected crime, and personal demographics) to detect patterns that it determines are predictive of recidivism based on the patterns learned from its training set.

Algorithmic Decision Systems

I use the term "Algorithmic Decision System" (ADS) to refer to automated, data-driven processes used for decision-making purposes. The level of sophistication of these ADSs varies widely, but they are often software-based. Technical approaches include simple rule-based systems, inferential statistical methods, and a variety of advanced machine learning methods (Selinger and Seager 2012, Yeung 2017). In some circumstances, this term may also be appropriate for actuarial processes that do not require software. In addition to the term "Algorithmic Decision System" (ADS), I will

also describe these systems as subsets of ADSs such as “decision-making systems” (i.e., a fully automated ADS) and “Decision Aids”/”Decision-Aiding Systems” (i.e., decision guidance systems); however, most examples discussed, due to current technical limitations, will fall within algorithmic decision aiding systems of varying strengths. Where appropriate, I will also use names of types of ADS systems, most often “risk assessments.” Each of these instruments utilizes different data and methods and may produce different types of outputs including numerical scores, specific recommendations, or visual indicators (e.g., heat maps that highlight geographical areas as shades of red, yellow, and green, ranging from highest to lowest risk area).

Predictive Risk Assessments

One of the earliest types of algorithmic procedures used in government were actuarial processes, which use mathematical rules based on probabilities related to risk. Actuarial processes were developed for use in the criminal justice system as early as the 1920s. However, these algorithmic aids have become increasingly more sophisticated with the increase in access to data, the acceleration in computer processing power, and the advancement in statistical modeling and machine learning processes.

This has made algorithmic policies particularly alluring in spaces with complex social problems that humans had often proved fallible in addressing. Furthermore, given the ability to scale software for significantly less money than increasing the workforce in a sector or agency, algorithmic decision systems were seen as a way to increase efficiency and benefit resource allocation (Mayson 2019). In particular, algorithmic

decision systems, often in the form of risk assessments, have rapidly been adopted within a variety of criminal justice contexts, including policing, criminal courts, diversion programs, and prisons (Werth 2019).

Algorithmic Interventions

I use the term "Algorithmic Interventions" to refer to the choice to use an algorithmic tool, such as decision systems, as a part of a public policy solution. Such a choice can often be at the expense of another policy tool which may have attempted to serve a similar purpose, or as a standalone add on to an existing piece of legislation. This frames the existence of ADSs as a choice to introduce an algorithm into a policy area, and by extension, introduce the attendant effects and impacts an algorithm may have. This term is used less often than the other terms discussed above primarily because much of the analysis presupposes an algorithmic intervention has already occurred, which is why the area is useful for analysis. Nonetheless, this term is important because it not only describes specific phenomena of interest, but it also characterizes the legislative process broadly, i.e., policy making in the U.S. is currently undergoing an algorithmic intervention that needs to be understood in its full scope and impact.

Discussion and Conclusions

This dissertation seeks to unpack the issue of the adoption of legislation that includes decision-making and aiding algorithm, collectively described as Algorithmic Decision Systems (ADS). In order to understand ADSs as a distinct kind of political tool

and the implications they may have on both policymaking and implementation, I analyze the following questions:

(1) What Motivates Policymakers to Support Algorithmic Decision Systems?

(2) Do the Perceived Benefits of Algorithmic Decision Systems Increase Policymaker Support for ADS-Inclusive Legislation?

(2a) Do fluctuations in the narrative about Algorithmic Decision Systems Alter Levels of Legislative Support?

Machine learning technologies may present a tremendous opportunity to leverage big data for social reform that is not influenced or corrupted by human bias, but using such emerging technologies before they are adequately designed with appropriate oversight and discussion among relevant stakeholders, may lead to unexpected social and institutional harms. However, the choice to use these decision systems is not made in a vacuum, but rather implemented within complex policy environments by actors with political and institutional motivations and constraints. My dissertation seeks to unpack the adoption of ADSs in order to better understand some of the causes and consequences of governmental adoption of such systems.

Chapter 3
Algorithmic Decision Systems as Policy Tools

Algorithmic Decision Systems are becoming deeply integrated in the institutional design of governing systems. Outside of full automation, these systems provide scores, recommendations, or other advisory information and are ultimately meant to inform, influence, or alter the discretion of human actors. For instance, this may be in the form of providing restrictions that the human actor must act within, such as when an algorithmic score set a discretionary floor or ceiling (for example, an output that suggests a criminal sentence no less than 5 years or no more than 5 years) within a set range of decisions. In a less rigid system, an algorithmic decision aid may primarily influence decisions through informational framing, such as when an actor is not required to incorporate the information provided by the algorithm, but still has seen it and may be cognitively "nudged" or influenced in some manner either consciously or subconsciously (Yeung 2017; also see Sunstein 2014). Some of the most well-known decision-aiding software are used within criminal justice. For instance, predictive policing software used to allocate policing resources based on a crime occurring in a certain location or being committed by a specific person.[23] However, many similar systems are used throughout other governing and public institutions, ranging from welfare distribution to child safety assessments. While this software is often procured solely at the discretion of the agency

[23] Geographical clusters, often illustrated using heat maps, are a more dominant framework for predictive policing software; however, other frameworks, such as Chicago's "Strategic Subject List (SSL), often referred to as Chicago's "heat list" have been used to identify threats on an individual-level.

or institution in which it will be used, increasingly, policymakers are considering including the mandatory adoption of these tools within legislation.

The following chapter argues that the inclusion of algorithmic decision aids in legislation is a burgeoning and unique policy tool that can enable decisionmakers to mitigate policy drift.[24] In particular, the flexibility and dynamic nature of algorithmic aids can supplement discretion bands or other traditional bureaucratic constraints. However, these same qualities also ensure a substantial level of political and policy flexibility in the implementation and use of these tools. To illustrate this, I explore the ways in which the legal rigidity of the algorithmic aid can be adjusted based on the policymaker's preferences for the outcome and their confidence in the end-actor's[25] willingness and ability to execute the spirit of the policy, as well as the ways that the end-actor may respond to those constraints. The intent of this analysis is to unpack and analyze the contexts in which algorithmic systems are adopted and the consequences of these choices utilizing a formal game-theoretic framework.

The use of a formal game-theoretic framework allows for the exploration of how a system, represented by the game-theoretic model, may change as aspects of the system are changed. This substantially advances two major goals of the dissertation: first, by analogizing the inclusion of ADSs in legislation to a situation of bureaucratic delegation by a legislator, I explicitly make the argument that ADSs are political artifacts in the

[24] Policy drift refers to policy change that occurs outside of formal legislative reform (Béland, Rocco, & Waddan 2016).

[25] In the context of algorithmic decision aids, the end-actor may be any actor who will ultimately execute a policy decision. This could be bureaucrats, law enforcement officials, judges, and other public sector employees.

same way that other types of constraints and grants of powers to bureaucracy, like rule making authority or menu laws, are inherently political artifacts. This is in line with the discussion from the prior chapter that argued that ADSs are a natural evolution of bureaucratic processes. Second, relying on a game-theoretic model allows for an explicit evaluation of the impact of ADSs on a part of the legislative process. For instance the following model explicitly unpacks how the of an ADS changes the equilibrium level of discretion and the level of legislator support for delegation to a bureaucrat, in relation to these equilibrium levels when an algorithmic decision system is not included. This focuses on understanding how use and adoption of ADSs has specific process impacts, in this specific example, by helping legislators curtail bureaucratic drift and alleviate the costs associated with the principal agent problem involved with delegating policy making authority to a bureaucrat. This model also ties into the broader theme of how ADSs may have broader systemic advantages, relative to other policy tools, that may make them may help them propagate through different policy areas.

Background

Governments have been integrating cyber technologies for more than half a century, but the social and technical advancement of these technologies has recently positioned them to expand beyond simple tools of efficiency and into key points of public policy. These tools are sometimes adopted by a specific agency or actor, for instance, by a police department, but they are also increasingly included in legislation, meaning that policymakers choose to require agencies to adopt these tools.

The choice to support or oppose the adoption of an ADS is not only based on a policymaker's independent policy goals, but also the expected behavior of others in the issue area's policy making space. Therefore, in a traditional spatial framework, the decision to include this software into legislation is also based on how this action may or may not move the outcome closer or farther away from the policymaker's ideal point, which is embedded in a complex multi-actor system. Algorithmic aids may encourage policymakers with differing goals to support the same legislation. This may occur due to the opacity and complexity of the technology. For example, these programs are often 1- proprietary, which limits access to understanding their inner workings (Pasquale 2015) and 2- the technology's highly complex attributes, which have rendered the rational or process for machine learning outputs uninterpretable to humans (Weinberger 2019). This confusing landscape is often veiled with the idea that Machine Learning (ML) outputs are mathematical truths and purely scientific. Thus, anyone who believes that their own rationale is objective and reflective of the truth may think that an "objective," "scientific" algorithm will produce the same conclusions that they have come to, for instance, that many defendants are dangerous and should be locked away or, conversely, that most defendants' crimes were highly contextual and that they do not pose a threat to society and should not be incarcerated. This may persuade policymakers on different sides of an issue to support a bill with algorithmic decision aids that will help to guide decisions toward the "right" outcome, which competing sides can rationalize as being parallel to their own beliefs.

A Legislative Perspective of Bureaucratic Discretion

A core foundation of modern representational democracies rests in their ability to accommodate the participation of a complex set of actors within the governing process. Thus, the primary decision makers in a democratic system must anticipate how their preferences will be executed if they want the outcome to align with their objectives. Many institutional, behavioral, and circumstantial dynamics may cause the outcome to differ from the decision makers' initial intent. To counter such policy drift, a decision maker may employ specific strategies or institutional tools. For example, legislators,[26] who must depend on varying levels within the bureaucracy to correctly interpret and effectively execute their legislation, must decide what discretionary choices to allow the bureaucrats, such as what aspects of a policy must be mandatory or otherwise difficult to avoid or alter.

One might think that for the legislator to ensure cooperation, they should make their policies as rigid as possible in order to deter bureaucrats from intentionally or unintentionally straying from the intent of the policy. However, prior studies have shown that, from the principal policymaker's standpoint, there is a delicate balance between political control and agency flexibility (McCubbins et al. 1987, Besley and McLaren 1993, Epstein and O'Halloran 1999, Bendor and Meirowitz 2004). More rigid policies, such as three-strike laws, may enable policymakers to more precisely dictate policy implementation, but this would also limit bureaucrats' ability to respond to new

[26] Many of these examples will be illustrated in the context of principal decisionmakers, such as policymakers/ legislators and subsequent agents, such as bureaucrats; however, the analysis is relevant to any decision-making principal who relies on agents to execute instructions.

information or changing circumstances on the ground which may threaten the ultimate success of the policy. It also neglects to take advantage of the substantive expertise of an agency or even dissuade competent, well-trained bureaucrats from service (Besley and McLaren 1993). Thus, policymakers must balance the desire for bureaucrats to dynamically adjust policies based on the bureaucrat's expertise, as well as their access to new or changing information, while also constraining the agents enough to ensure that they act in accordance to the spirit of the policy. Deviation from the original intention of the policy may be a consequence of politically or ideologically motivated bureaucrats who may intentionally sabotage or subvert the execution of the policy (Gailmard 2002), the ability for the agents to understand and interpret both the intention of the policy and how to best achieve its ends (Bendor and Meirowitz 2004), or to the institution's capacity which may not give the agency the needed resources to effectively implement the policy (Huber and McCarty 2004). Each of these circumstances requires different actions to successfully overcome the issue and achieve the policymaker's desired outcome.

Strategic legislators may attempt to account for these competing challenges when crafting legislation. If the policymaker thinks an agency may stray from direct implementation of the policy at hand, they then must choose which policy tools will restrain those agencies while balancing the harms that may come from restricting the agency's responsiveness to new information or circumstances. Traditional policy tools include direct alterations of an agent's discretionary power. For example, policymakers may limit discretion by creating highly rigid procedures, such as mandatory minimums. More moderate restrictions may be in the form of "menu laws," which offer bureaucrats

predefined options to choose from or "discretion windows" which allow some choice but only within specified bounds (Gailmard 2009). More discretionary power may come in the form of open or loosely defined directives. Policymakers must choose which tools will best serve their long-term policy objectives, while also balancing the threat of opposing decision makers rejecting the entirety of the policy. Such opposition may include an executive veto power (Epstein and O'Halloran 1999; Volden 2002), external pressures such as lobbying interests (McCubbins et al. 1987), or competing legislative actors or groups who oppose the legislation or cause intra-coalitional conflicts (Horn and Shepsle, 1989).

Algorithmic decision systems used as a policy tool.

Algorithmic decision aids provide policymakers with a new policy tool to use when mediating the tradeoffs between control and context. These algorithmic decision aids are novel policy tools, in part, because they are a constraining structure that has predefined outcomes prescribed by non-policymakers (e.g., software engineers), but also in that they could be designed to dynamically change based on real-time input using machine learning or changed or revised by being manually updated or retrained. Dynamically updating systems refer to machine learning techniques that adjust or "learn" from past data to iteratively update predicted outcomes to better fit the circumstances, while systems that are manually updated refer to static applications that only change when their program is manually entered and altered. Either of these approaches would make the tool a more flexible constraint than traditional policies, such as predetermined

discretionary windows changes to which would require new legislation. This is not to necessarily say that these tools are a better policy choice or to argue that the flexibility or adaptability of these tools deliver on their promises. However, it is to claim that the way that decision makers view these tools will have an impact on their adoption during the policy process.

The adoption of algorithmic decision-making or aiding systems may also be paired with further flexibility via the level of legal or procedural authority[27] that the policy gives the software. A more rigid system could mandate or otherwise require an agent to make decisions that are based on the algorithmic assessment. For instance, a criminal recidivism risk assessment score of 10 out of 10 at a sentencing hearing could require a minimum sentence length of 6 months. Conversely, a less rigid system could be deployed as an optional decision aid with no authority. This aid can then be used at the agent's will as supplementary information, such as only providing the judge with a risk assessment score without an explicit recommendation so that it may be used as supplementary case information.

There are various moderating choices between these opposing ends of the spectrum, such as requiring only some algorithmic outcomes to be binding decisions so that more extreme outcomes would have a mandated action associated with them, but in the case of more moderate outcomes, actions were left up to the agent. For instance, a risk assessment of 9 or higher may require a judge to deny bail, while a score of 5 may be

[27] Legal rigidity refers to a requirement that the algorithm's output has an influence on the decision, while procedural rigidity refers to when the algorithm's output is not required to be considered or used, but is due to reasons like a de facto trust in its assessment.

left up to the judge's discretion. This is the case for a child welfare risk assessment used by Allegany County, PA. When suspected abuse or neglect is reported to a hotline, a volunteer is also given a score that attempts to predict the likelihood of the child being in significant danger. The system scores the risk from 1 (very low risk) to 20 (very high-risk). Most the time, this system is only meant to provide the volunteer with further information before deciding whether to open a case which will go to a social worker for investigation. However, in cases that score a 20, the highest score, a case is automatically opened (Chouldechova et al. 2018). It is also important to consider that even an algorithm with less authority may produce de facto adherence to the software's output or suggestion. For instance, in the Allegany County case, many volunteers said that they often reconsider their own determination if the algorithm provides a significantly different score than they had expected. This is due to their trust in the software being objective and effective (Eubanks 2018). This trend may produce unintended consequences if legislators are unaware of the effect that scorings and advising systems may have even when they are very weak. Both the use of the algorithm and the amount of authority it is given provide legislators with further tools to balance the tradeoffs of delegation.

The Model

Previous models of delegation decisions have unpacked the complex stages of decision-making and execution in a range of scenarios. Many of these studies have produced basic principles with which scholars, legislators, or other interested parties can

use to anticipate organizational behavior and policy outcomes. Researchers often focus on the ways in which different policy tools and procedures are effectively used by legislators to mitigate bureaucratic drift and strengthen political control. The proposed model seeks to include algorithmic decision aids within this discussion as a unique policy tool that can enable decision principals to mitigate agent drift that 1) create discretionary bands of varying strength and legal rigidity and 2) enable more flexible and dynamic moderation than traditional bureaucratic constraints.

This section proceeds by discussing the general form of a delegation model, utilizing the canonical Epstein and O'Halloran 1996 model of an interaction between Congress (C) and an executive agency (A). This model is generalized to any legislature (L) and any bureaucrat (B), who are engaged in a delegatory relationship. L moves first by setting a policy (p) with some discretionary band (d), nature then chooses a policy shock (ω) that is observed by B, who sets the ultimate policy outcome (x). Both L and B value x in accordance with their policy preferences, i.e., x is valued based on how small the difference between it and the ideal point of L and B, respectively. This model is then expanded in two essential ways: 1) a dynamic time component is added to indicate that a delegation game is observed over multiple rounds where the legislature only moves once at the beginning and the bureaucrat utilizes its discretion to respond to different shocks chosen by nature over time, and 2) by adding a constraining algorithm (a) that also changes over time, which is used by the legislature to dynamically limit the actual use of the bureaucrat's full discretionary authority. The object is to understand how inclusion of an algorithmic component changes the optimal level of discretionary authority (d*)

chosen by the Legislature. This is done by comparing the equilibrium level of discretionary authority (d*) under the original Epstein and O'Halloran and the modified algorithmic model.

The following models build on Epstein and O'Halloran's 1996 model of legislative delegation and agency discretion. The Epstein and O'Halloran model starts with a realized outcome:

$x = p + \omega + d$, where

x is the final policy outcome that is executed in the world,

p is the policy chosen by an agency,

ω represents exogenous shocks that may affect the policy and shape the outcome, and

d is the discretion given to the agency

Traditionally, legislators (or other decision principals) are incentivized to give the agency discretion because their expertise and proximity to the issue will mean that only the agent will be able to observe ω. In order to balance the tradeoffs of political control and bureaucratic expertise, the legislature can set the level of discretion (d) on an interval $[0, \infty)$, the agency can then choose the specific policy parameters (p), given the constraint that $p \in [l, r]$ (Bendor and Meirowitz 2004; Gehlbach 2013).

I expand this baseline model to illustrate how algorithmic decision aids are used within the policy process by adding a as an algorithmic policy moderator on agency discretion, where a is any value between 1 (no algorithmic aid) and 0 (a fully binding algorithm).In this sense, a functions as the output of an algorithmic process in the form of

a constraint. The algorithm dynamically evaluates a set of data and chooses an appropriate proportion of an agency's full discretion that ought to be used to set the policy before ω is actually observed. Thus, an algorithmic output of $a = 1$ would provide no constraint on the discretion given to an agency, allowing them full use of their delegated authority, d. This equation takes the form of

$x = p + \omega + da$, where

x is the final policy outcome that is executed in the world,

p is the policy chosen by an agency,

ω represents exogenous shocks that may affect the policy and shape the outcome,

d is the discretion given to the agency, and

a is the constraint level, bounded on the interval [0,1] chosen by the algorithm[28]

By implementing an algorithmic aid, policymakers can both put bounds on discretion as well as absorb some amount of exogenous shocks.

Prior models have found that greater policy uncertainty also increases the cost, limiting an agency's ability to absorb ω. However, the dynamic and flexible nature of the algorithmic system may give the principal a way to limit discretion without the full cost

[28] In the original Epstein and O'Halloran model, d is set on the interval [*l,r*]; however, introducing a moderating factor, *a*, creates a distinction between the policy level of *d* set by the Legislator in the initial authority granting piece of legislation, and the effective discretionary level, *da*, which is the actual amount of discretion that the Bureaucrat can use to move policy and set the final outcome policy *x*. In this sense, for the current model, the policy level of *d* is bounded on the interval $[0, \infty)$: the Legislator could conceivably grant any level of discretionary authority to the Bureaucrat; however, as will be demonstrated later in the model, this is done with respect to the expected algorithmic constraint, *a*, such that the effective discretionary level, *da*, is still bounded on the interval [l,r].

of unknown exogenous shocks. This is because algorithmic aids, through either machine learning processes or manual updates, can be adjusted to exogenous shocks that would also affect the level of discretion conveyed to agencies. To represent the fundamental learning process over time that the algorithm engages in, in addition to including the notion that bureaucratic discretion and policy setting occurs over time, the model incorporates time in a dynamic process.

In this model, $x_t = p + \omega_t + da_{t-1}$, where t represents time. To reiterate, x is the realized policy implemented in the world after the legislature sets a core policy (p), nature chooses a shock (ω_t), and the bureaucrat applies its algorithmically constrained level of discretion (da_{t-1}). The main difference between this setup and the Epstein and O'Halloran model is that bureaucratic policy setting happens at multiple points in time, as opposed to only once after the initial law and initial shock are observed. However, the timing of events is also slightly different between the two models. In the current model, the timing is as follows:

> (1) a legislature (L), sets a policy (p), a discretion level (d), and authorizes the use of a constraining algorithm (a);
>
> (2) the algorithm(a) is trained on prior history and processes that are identified as important to a policy area;
>
> (3) the algorithm chooses to constrain the full bureaucratic discretion of an agency by some proportion based on prior history and processes(a_{t-1});
>
> (4) an exogenous shock (ω_t) is realized and evaluated by the bureaucratic agency;

(5) the agency uses is mediated discretion level (da_{t-1}) to implement a policy (x_t);

(6) steps 2-5 are repeated.

The outcome and timing changes introduced by the model require a reevaluation of the utility functions from the original E&H model. Their initial work observed the following utility functions for Congress (C) and an agency (A):

$U_C = -(C - x)^2$ with C being Congress' ideal point and x being the outcome policy

Assume that $C = 0$, so $U_C = -x^2$

$U_A = -(A - x)^2$ with A being the Agency's ideal point and x being the outcome policy[29].

Following this setup, but generalizing to a legislator (L) and bureaucrat (B)[30], yields the following utilities:

$$U_L = -(L - x)^2 = -x^2$$
$$U_B = -(B - x)^2$$

However, since the algorithmic aid, a_{t-1}, would change over time, along with the shocks(ω), bureaucratic policy, and final outcome policy, the utility evaluations would also change over time. This can be illustrated as

$$-x_1^2 + \delta(-x_2^2) + d^2(-x_3^2) \dots \delta^{n-1}(-x_n^2)$$

[29] Both C and A are constrained to be on the policy space *[l,r]*.
[30] By extension, L and B are also constrained to be on the policy space *[l,r]*

with δ being the discount rate in which $\delta \in (0,1)$. This pattern can be simplified into summation notation, $\sum_{t=1}^{\infty} \delta^{t-1}(-x_t^2)$. The quantity x_t, the realized policy at time *t*, would be predicated on the information available when the algorithm was updated. However, the legislature does not make its choice of policy at every point in time, instead delegating to the bureaucrat and the algorithm. Under the assumption that much of the purpose of this delegation is to avoid large shifts in outcomes x and induce policy stability, then the best position for the legislator (L) is to assume that the combined discretionary authority and algorithm work to stabilize x_t to some expected equilibrium value x. While the realized value of x would change over time, thus reverting to x_t for the bureaucrat, the best information that the legislator has to work with is to assume that a stable policy is reached on average, in order to effectively evaluate the utility of any proposed strategy. Taking this information into account, the summation changes to

$$\sum_{t=1}^{\infty} \delta^{t-1}(-x^2),$$

which is a straightforward sum of an infinite geometric series. This results in a value of $-\frac{x^2}{1-\delta}$, so that $U_L = -\frac{x^2}{1-\delta_L}$.

However, to simplify some of the math, I have altered the Epstein and O'Halloran model to include a shape parameter of $\frac{3}{2}$ to the legislator's utility function, instead of the $\frac{1}{2R}$ that can be found in the expected utility function of the original model.[31] This does not change the underlying relationship of the parameters within the model, but rather, the

[31] R represents the right endpoint of a policy shock space, i.e. a continuum of potential values that a policy shock, like a change in unemployment, a natural disaster, or a security incident, could take. R will be explicitly defined and discussed later in the model.

sensitivity of the utility to changes in the parameters, which does not contribute any analytical substance to the point at hand. This results in Equation 1.

$$U_L(x) = -\frac{3x^2}{2(1-\delta)} \quad [1]$$

A legislator must base their decisions on a longer time horizon, but a bureaucrat will interact with the policy in independent cases over time. Legislators make a singular policy that must fit a broad range of cases over time, but a bureaucrat will decide when and how to implement a policy at the time of each applicable case. To illustrate this, I have expanded the model to show the legislator's utility includes the infinite payoff stream while the bureaucrat's does not. Equation 2 illustrates the difference in the bureaucrat's decision-making process.

$$U_B(x) = -(B - x_t)^2 \quad [2]$$

To analyze the bureaucrat's decision at each point in time, the equation below illustrates that the agency will choose the outcome closest to its ideal point, given its level of effective discretion (da_{t-1}). Specifically, the first line illustrates when given a shock, ($-\omega$), that is farther away from their ideal point than their discretion band covers, the bureaucrat will move policy to the right as far as their effective discretionary band allows. The second line shows that when a shock is sufficiently close to the bureaucrat's ideal point such that it lies within the discretion band around their ideal point, the bureaucrat will set the outcome policy to their ideal point (B). The third line shows that when a shock, (ω), lies to the right of the discretion band around the bureaucrat's ideal

point, the bureaucrat will move policy to the left as much as its effective discretion allows.

$$x_t = \begin{cases} \omega_t + SQ + da_{t-1} & if\ l \leq \omega_t < B - SQ - da_{t-1} \\ B & if\ B - SQ - da_{t-1} \leq \omega_t \leq B - SQ + da_{t-1} \\ \omega_t + SQ - da_{t-1} & if\ B - SQ + da_{t-1} < \omega_t \leq r \end{cases} \quad [3]$$

The model relies on a fairly standard Nash equilibrium solution concept. Utilities and preferences for each actor are common knowledge, so the goal is to produce a solution that is a best response for each actor in the model. Since the timing of the game is well structured and common knowledge, the Legislator will choose a best response function that maximizes their utility first, which involves setting the optimal level of discretionary authority, d*. After this, the Bureaucrat will choose a best response function that maximizes their utility subject to d=d*, i.e., that the Legislator has chosen a best response. In this way, both actors are responding to each other with their respective best response functions, creating a standard Nash Equilibrium. Using backward induction to solve for the best response functions, the last move of the game is the bureaucrat's decision to set the policy at the time of implementation. Knowing that the bureaucrat's decision is analogous to Equation 3, the legislature will set an optimal Status Quo (SQ^*) and d (d^*) to maximize their expected utility on the move prior. This can be modeled by evaluating the legislator's utility function for each of the separate outcomes at each of the separate intervals that govern the bureaucrat's decision illustrated in Equation 3.

Integrating over these separate evaluations yields the total expected utility for the legislator as seen in Equation 4.

$$EU_L = -\frac{3}{2}\int_{l}^{B-SQ-da}\frac{(\omega+SQ+da)^2}{1-\delta}\,d\omega - \int_{B-SQ-da}^{B-SQ+da}\frac{B^2}{1-\delta}\,d\omega - \frac{3}{2}\int_{B-SQ+da}^{r}\frac{(\omega+SQ-da)^2}{1-\delta}\,d\omega \quad [4]$$

Step 1: Explicitly evaluate the integrals by substituting the interval endpoints for ω in the antiderivative of the legislature's utility function. This allows for the evaluation of each possible realized policy outcome over the interval for which that outcome is possible.

$$= -\frac{3}{2(1-\delta)}\left[\frac{B^3}{3} - \frac{(l+SQ+da)^3}{3}\right] - \frac{3}{2(1-\delta)}\left[\frac{3B^2(B-SQ+da-B+SQ+da)}{3}\right] - \frac{3}{2(1-\delta)}\left[\frac{(r+SQ-da)^3}{3} - \frac{B^3}{3}\right] \quad [5]$$

Step 2: Factor out and cancel the constants. Equation 5 gives the total utility for each possible policy outcome over the interval where that outcome is possible; by summing over these distinct possibilities, I obtain the total utility for the Legislator over the policy space (from *l* to *r*).

$$= -\frac{1}{2(1-\delta)}[B^3 - (l+SQ+da)^3 + 6B^2da + (r+SQ-da)^3 - B^3] \quad [6]$$

Step 3: Simplify. Equation 6 represents the full evaluation of the Legislator's utility function, but it is analytically complex. However, by distributing the multiplicative term and canceling out the B^3 terms, I obtain a slightly simpler equation for the evaluation of the Legislator's utility function.

$$= -\frac{(r + SQ - da)^3 - (l + SQ + da)^3 + 6B^2 da}{2(1-\delta)} \quad [7]$$

Step 4: Expand the cubic expressions. Equation 7 is still too complex for straightforward evaluation of how the Legislator's utility changes with respect to the level of discretion they can set. To move forward, the cubic expressions are expanded to better isolate each term in the equation.

$$= \frac{2(da)^3 - 3r(da)^2 + 3l(da)^2 + 3r^2 da + 6rSQda + 6SQ^2 da + 3l^2 da + 6lSQda - 6B^2 da + l^3 + 3l^2 SQ + 3lSQ^2 - r^3 - 3r^2 SQ - 3rSQ^2}{2(1-\delta)} \quad [8]$$

Equation 8 is the final form of the Legislator's utility function evaluated over the entirety of the policy space (*l,r)*; however, given the complexity, this equation is analytically intractable. To obtain a more analytically manageable equilibrium outcome some additional assumptions are necessary. The current model assumes that the endpoints of the shock interval, $[l, r]$, only relate to each other in that $l < r$. However, under the assumption, as in Epstein and O'Halloran, that the shock interval is symmetric about the legislature's ideal point, then the interval on which shocks exist becomes $[-R, R]$, where I have substituted $-R\ for\ l$ and $R\ for\ r$. This results in Equation 9.

$$EU_L = \frac{(da)^3 - 3R(da)^2 + 3R^2 da + 3SQ^2 da - 3B^2 da + R^3 + 3RSQ^2}{1-\delta} \quad [9]$$

From here, I explicitly evaluate the values of SQ and d that maximize the legislator's expected utility (EU_L^*) from delegation. I begin with SQ in Equation 10:

$$\frac{\partial EU_L}{\partial SQ} = \frac{6SQda + 6RSQ}{1-\delta} = 0 \quad [10]$$

which results in $SQ^* = 0$. Practically, this means that the Legislator sets the existing policy to be their ideal point.

The next step is to solve the optimization problem for the legislator's expected utility function with respect to d, which will provide two critical points for potential values of d* or the optimal level of discretion in equilibrium. The critical points are shown in Equation 11:

$$\frac{\partial EU_L}{\partial d} = \frac{3a^3d^2 - 6Ra^2d + 3R^2a - 3B^2a}{1-\delta} = 0 \quad [11]$$
$$= \frac{-3a(B + R - da)(B - R + da)}{1-\delta} = 0$$
$$d = \frac{R+B}{a}, \frac{R-B}{a}$$

Equation 11 shows that there are two potential solutions for the discretion level, d, that would optimize the legislator's expected utility function. Determining which value in Equation 11 maximizes the Legislator's expected utility requires a second derivative test to determine which value of d has a negative second derivative value, as this will be the value of d that maximizes the expected utility function, or d*. The second derivative test is shown below for both potential values of d* in Equation 11.

$$\frac{\partial^2 EU_L}{\partial d^2} = \frac{6a^3d - 6Ra^2}{1-\delta} = \frac{6a^2(ad - R)}{1-\delta}$$

$$\frac{6a^2(a\left[\frac{R+B}{a}\right] - R)}{1-\delta}, \frac{6a^2(a\left[\frac{R-B}{a}\right] - R)}{1-\delta}$$

$$\frac{6a^2B}{1-\delta}, -\frac{6a^2B}{1-\delta}$$

The model considers the situation where the Bureaucrat is more conservative than the Legislator. Since the Legislator's ideal point is fixed at 0, the Bureaucrat's ideal point, B, must be to the right of the Legislator's ideal point, or B > 0. Likewise, δ is a discount factor bounded on the interval (0,1). This means that all the variables in the second derivative are positive, so the only mathematically valid solution for the maximum, i.e. where the second derivative is negative, is $\frac{R-B}{a}$. Thus the value of d that maximizes the Legislator's expected utility is formally stated below in Equation 12[32].

$$d^* = \frac{R-B}{a} \qquad [12]$$

Discussion

The results of the above model indicate that algorithmic interventions in the policy process produce new insights that change our understanding of bureaucratic delegation. I will discuss several analytic conclusions drawing from the results of Equation 12: 1) the equilibrium discretion outcome delegated by the legislature at the margins, i.e., the endpoints of algorithmic constraints; 2) the positive relationship

[32] Equation 12 also indicates that the effective discretion level, *d*a,* does not exceed the bounds of the policy space, *R*. This is always true given that $d^*a = R - B \leq R$.

between bureaucratic capacity and algorithms, and 3) the inverse relationship between algorithmic constraint and discretionary authority.

The initial finding concerns the equilibrium level of discretionary authority at the margins of the output of a constraining algorithm. Given that $a \in (0,1]$[33], a value of 1 indicates that there is no algorithmic constraint, while a value of 0 indicates that there is no human discretion. Analytically, a value of 0 would be the result when the algorithm has complete authority, which would functionally, if not actually, mean that the policy area was completely automated. We can consider the case when $a = 0$, by explicitly evaluating the optimization problem with equation 9 at that value. This yields equation 13,

$$\frac{\partial EU_{L \mid a=0}}{\partial d} = \frac{\partial}{\partial d}\frac{R^3 + 3RSQ^2}{1-\delta} = 0 \qquad [13]$$

Since d is no longer an explicit term in the optimization problem, the partial derivative reduces to 0. This yields the equality $0 = 0$. Analytically, this indicates that there is no value of d which does not satisfy the optimization condition. In other words, the Legislator is comfortable giving a Bureaucrat any level of discretion if the algorithm is

[33] As discussed in footnote 28, d, the level of bureaucratic discretion also exists on a strictly positive interval. It can minimally take on the value of 0, where the Legislator does not delegate any authority to the Bureaucrat. In principal, the upper limit of d is unbounded: the Legislator can delegate as much authority to move policy outcomes as they wish; however, within the confines of the model, and practically speaking, a Bureaucrat would only exercise as much discretion as necessary to move a policy to their ideal point. At maximum, this would require enough discretion for a Bureaucrat whose ideal point was at one end of the policy shock space, say –R, to be able to move the final outcome policy to that ideal point from the opposite end of the shock space, R. This yields a maximal value of 2R. However, the practical amount of discretionary authority utilized by the Bureaucrat is not reflected by just *d*, but is also impacted by *a*, the algorithmic constraint. This results in a final, effective discretionary authority level being bounded on the interval $da \in [0, 2R]$.

functionally autonomously setting the final policy. This makes a certain amount of sense, since in an autonomous setting, bureaucratic discretion is moot, so there is no level of discretion which would change how the Legislator views the situation.[34]

The results when $a = 1$ are slightly more interesting because they are functionally equivalent to the equilibrium discretion level in the Epstein and O'Halloran model $d^* = R - P$, where P is the ideal point of the President. This validates the model in part, since an algorithmic constraint value of 1 means that the algorithm is allowing the bureaucrat to exercise complete discretionary authority, making the algorithm essentially non-existent for that period. Thus, when the algorithm is not playing a role, the results of the model are the same as a model without an algorithmic constraint. This emphasizes the need to understand if claims of a suboptimal performance by an algorithmic decision aid is because of the decision system itself or if it because it is not being used.

There is an additional, and in this instance somewhat unexpected, finding that can be derived by comparing the model's equilibrium discretionary outcome to that obtained by the original model. Since $a \in (0,1]$, for all instances $a \neq 1$, the utility for the legislature adopting an algorithmic constraint policy is at least as high as a policy under which no algorithmic constraint is adopted. For a fixed level of utility to a legislature, comparing two policy schemes, one with an algorithmic constraint and one without, the

[34] A close reading of the model shows that the algorithm does not have any ability to unilaterally alter policy outcomes; by constraining the Bureaucrat to not exercise any discretion, the outcome policy cannot be altered after the policy shock is observed. Within the confines of the model, this is preferable to the Legislator if the policy shock consistently pushes the policy to a value between the Legislator's and the Bureaucrat's ideal point, i.e. when the Bureaucrat would consistently utilize their discretion to move policy away from the Legislator's ideal point.

legislator would be more willing to extend discretionary authority, in equilibrium, when an algorithmic constraint is part of the policy. Functionally, under the assumption that increases in discretionary authority are analogous to increases in bureaucratic capacity (i.e., to move policy to a larger degree, a bureaucratic agency must have more capacity, resources, expertise, etc. to utilize the full extent of its discretionary authority) then including an algorithmic constraint should increase bureaucratic capacity, in equilibrium. Huber and McCarty (2004) argue that if policymakers believe an agency will not be able to accurately execute a policy, for any reason including insubstantial resources, they are likely to recoil pursuing or supporting legislation in this area.

Finally, the model suggests that legislators may hamper bureaucrats' ability to respond to shocks in contexts where algorithmic aids poorly understand the nature of the exogenous shock[35], such as in situations marked by high complexity, or low data quality. This can be seen by taking the partial derivative of the equilibrium discretion level with respect to the algorithmic constraint: $\frac{\partial d^*}{\partial a} = \frac{B-R}{a^2}$. Since R is the upper bound of the realized shock space, and B is constrained to be strictly less than R, the quantity $B - R$ is strictly negative, while a^2 will always be a positive value, making the partial derivative

[35] The nature of the relation between the underlying exogenous shock generation process and the algorithmic constraint output is not explicitly defined here. But for the purposes of the discussion, the algorithm fundamentally tries to predict a specific t | t-1. Where the prior distribution of t is well defined, understood, and predictable, the algorithm will produce a more confident estimate of t. The distance of this t from the status quo policy and the confidence of the algorithm in this estimate jointly determine the algorithmic constraint parameter. To the extent that the algorithm is consistently not confident in its estimates of t, then regardless of the distance of the estimate from the status quo, the algorithm will return higher levels of a.

monotonically negative. Thus, any increase in the algorithmic constraint parameter, a, is associated with decreases in the discretionary authority d. As a result, higher values of a, where the algorithm only slightly constrains the bureaucrat, are associated with smaller levels of bureaucratic discretion. Since the algorithm does not increase bureaucratic discretion, it only constrains it, adopting a policy that limits authority relative to an algorithm that would more reliably constrain the bureaucrat, only leaves the agency less able to deal with extreme shocks, albeit with full, if diminished, discretionary authority. These situations may arise in various governmental environments due to the many multifaceted social issues that have been difficult to understand or quantify. This could also occur in contexts that just do not have enough data to accurately predict or adjust outcomes for exogenous shocks; this could occur for several reasons. For example, predictive algorithms cannot learn from data that is not collected, for regulatory or political purposes; researchers may not have a thorough understanding of the relevant inputs to train algorithms on, or policy areas may be described by rare events and are correspondingly difficult to interpret accurately or predict.

Each of these findings provides reasons to reevaluate our understanding and expectations of bureaucratic delegation processes under the addition of algorithmic aids. Given the fact that algorithms, machine learning, and AI are being readily adopted in areas well beyond bureaucratic delegation, we must learn to appropriately situate this new component within our broader understanding of political and governmental processes.

Conclusion

The model and results presented in this chapter provides a clear rationale for why policymakers may be drawn to including an algorithmic decision system in a piece of legislation. Legislators are plagued by a consistent problem when faced with the choice of delegating authority to a bureaucrat. On the one hand, legislators do not have as much relevant subject matter expertise as a bureaucratic agent, making them less able to respond to on the ground conditions and shocks. But, extending the bureaucrat more discretion may allow them to pull the ultimate realized policy outcome away from the legislators ideal point and towards the bureaucrat. This principal-agent problem is at the core of delegatory interactions between legislators and bureaucrats.

Contributions and Novel Insights

By dynamically constraining bureaucrats' level of discretion based on contextual factors, ADSs offer legislators a partial solution. In instances where contexts are rapidly changing or that are subject to sudden shocks, for example financial market regulation or disaster response management, algorithmic policies promise to give bureaucrats full discretion when they need it, and constrained discretion when they do not, which in turn allows legislators to grant wider maximum discretion than they otherwise would. From the legislator's point of view the promise of an ADS is to provide the necessary level of discretion when needed, but to otherwise constrain the ability of a bureaucrat to shift a policy away from the legislators preference.

Of course, the lived reality of ADS usage may differ significantly. The performance, ability to constrain, and responsiveness of these systems are highly dependent on the actual context, including computational, design, and political contexts. However, the main point here is that ADSs, as a policy tool, offer potential systemic advantages, in theory, over similar policy tools. A legislator simultaneously concerned with granting wide latitude to a bureaucrat resulting in bureaucratic drift but also with overly constraining a bureaucrat and thus decreasing their ability to integrate useful information or respond to a shock sees a ready answer in ADSs: let the algorithm handle it.

The conceptual framework laid out by the model establishes the fundamental political process impacts that algorithmic policies can have. Further than that though, by analogizing ADSs to the situation of bureaucratic politics, I demonstrate how well such systems fit into the broader research program concerned with bureaucracy, providing evidence for the discussion in Chapter 2 of ADSs as a natural evolution of bureaucratic capacity and practice. By arguing that ADSs are fundamentally of a kind with broader bureaucratic policy tools and concerns, I also further the argument that ADSs are distinctly and deeply political phenomena that should be analyzed with the tools and methods of political science.

Future Research and Empirical Tests

As a concluding note, while an empirical evaluation of the formal model derived here is beyond the scope of this project, it is still useful to discuss ways the model may be evaluated empirically. Given the central contention that the model provides support for the notion that ADSs have structural advantages over similar policy tools, the findings here could be tested by examining empirical evidence for the robustness of such policy designs. Specifically, legislation that includes ADSs should be more likely to be reintroduced in subsequent sessions of a legislature, if they provide such systemic advantages. To the extent that ADSs are primed to propagate through policy areas, then the introduction of an ADS in one policy area should be associated with a higher likelihood of introduction in another policy area in a subsequent session of a legislature. In this way, ADSs may have diffusion effects that can be empirically observed. These empirical extensions are not the focus here; instead, the next chapter utilizes an empirical analysis to test for a different potential systemic benefit of ADSs than is tested by the model here: coalitional benefits for legislation that includes algorithmic policies. However, the empirical analysis is also concerned with understanding how the critical turn and resultant backlash phenomena, discussed in Chapters 1 and 2, may mitigate potential systemic advantages. In this way, the empirical analysis in Chapter 4 offers a more complete discussion of the main themes and goals of the dissertation than is discussed here: ADSs as political artifacts that have fundamental process impacts but that also situates the potential systemic advantages in relation to observed public scrutiny.

Chapter 4
Pretrial Risk Assessments and Criminal Justice Reform

Decision aids, in this case, pretrial risk assessments,[36] substantially increased in the 2000s and early 2010s, though after a recent backlash, have become much more politically fraught. Pretrial risk assessments, as a type of algorithmic decision system, represent one of the most fruitful instances of ADSs to test for potential systemic advantages and political process impacts, as well as to gauge how such potential advantages are moderated by backlash effects. This is due to the fact that pretrial risk assessments are one of the most visible forms of ADSs. Pretrial risk assessments have formed the focal point of many analyses, both popular and scholarly, that focus on the use of algorithmic policies. Pretrial risk assessments also represent one of the first, and most sustained, areas to be the focus of the critical turn and eventual backlash against ADSs. As such, this area represents one of the most promising policy tools to be able to test how ADSs might influence legislative processes, as well as how a backlash might moderate such process impacts. To test if the inclusion of an algorithmic decision aid provides advantages during the legislative process, such as increased bipartisan support, this chapter analyzes pretrial reform legislation in all 50 states in 2018, and all pretrial legislation that passed in a state legislature between 2012-2017. This timeframe will also be used to examine whether the changing narrative around risk assessments has

[36] Decision aiding algorithms in the context of pretrial decision aids are referred to risk assessment algorithms, so a variation of those terms will be used throughout.

corresponded with a decline in legislative support. This chapter will begin by discussing the pretrial process, reform efforts, and the adoption trends of pretrial risk assessments. The next section tests whether the inclusion of pretrial risk assessments in legislation results in structural advantages, such as increased bipartisan support, and whether or not this phenomenon changes over time.

This chapter substantively advances all three primary goals of the dissertation. The objective of this chapter is to provide initial empirical evidence that ADSs, in the form of pretrial risk assessments, have potential systemic advantages, and by extension political process impacts, on the legislative process (as specifically tested in the area of state level criminal justice legislation). This analysis directly advances the goals to provide a more empirical basis for research on algorithms and society as well as to specifically identify process impacts. Moreover, by directly positioning inclusion of algorithmic components as an independent variable in an analysis centered distinctly political dependent variables such as bipartisan support for legislation, this analysis furthers the argument that ADSs are inherently political. Beyond that, however, by integrating the effects of a potential backlash against such systems, this analysis demonstrates that ADSs are not just political objects themselves, but are the locus of distinctly political activity.

Case Selection: Pretrial Reform and Risk Assessment Algorithms

Over 10 million people are arrested each year in the United States (Federal Bureau of Investigations, 2017). After an arrest,[37] a pretrial process begins in which a judge or court magistrate must decide whether or not a defendant should be detained prior to their trial. Pretrial release decisions are primarily based on two factors in most jurisdictions: flight risk (i.e., the likelihood of failing to appear at the trial) and risk the defendant poses to individuals or the community (Kehl, Guo, and Kessler 2017). Specific procedures and laws vary by jurisdiction, but generally, judges may choose to 1) release the defendant on their own recognizance,[38] 2) release the defendant with conditions,[39] 3) set a monetary bail amount that the defendant must pay the court in order to be released and will forfeit if they fail to appear, or 4) detain the defendant without bail[40] (Feeley 2017).

The primary considerations for whether or not to release a defendant (and if so, on what conditions) tends to be the severity of the accused crime and the defendant's criminal background; however, judicial discretion enables judges and magistrates to consider other factors and contexts. These factors may be explicitly taken into consideration or they may implicitly contribute to a judge's intuition about a case (Guthrie, Rachlinski, and Wistrich 2007). Research indicates that judges' assessments

[37] While most procedures vary to some extent by jurisdiction, this is typically the process that occurs after an arrest or after a defendant turns them self into police following a warrant being issued. If a crime is reported, but officials chose not to issue a warrant, a summons is issued calling for the defendant to appear in court for either a pretrial appearance (at which this process begins) or at the trial.
[38] Release on recognizance only requires the defendant to sign a document stating they will return for their trial.
[39] Release conditions vary, but can include requirements like abstaining from drugs or alcohol.
[40] See Cohen and Austin 2008.

may include extralegal factors. For instance, a judge's may be influenced by their perception of the defendant's sociodemographic profile, such as the defendant's race, gender, or income. Arnold, Dobbie, and Yang (2018) find substantial bias against black defendants. Specifically, the authors find that judges, particularly less experienced judges, tend to set higher bail requirements for black defendants than for white defendants with similar crimes and criminal background. They argue this result likely stems from implicit biased regarding risks rather than explicit racial animosity. Defendants of lower economic status also face significant disadvantages, as well as institutional inequalities, such as having a public defense, rather than private counsel. Typically, public defense attorneys do not get to speak with their client until later in the process and often have more cases and fewer resources than private defense. Hissong and Wheeler (2017) find that bail tends to be at least 25% lower for defendants with private counsel. Furthermore, the authors find that 90% of the defendants with private counsel were able to post bail, in contrast to only 27% of those without private counsel.

Monetary bail compounds these biases because it is a fundamentally economically biased tool. Defendants are often given a bail amount that they must pay in order to be released and must forfeit the payment if they do not return for their trial (in addition to having a warrant issued for their arrest) (Rabuy and Kopf 2016). If a defendant cannot pay their bail in full, they may use a bail bond service. Bond agents typically require the defendant to pay 10% of the bail upfront, which will act as payment for the service and will not be returned. Additionally, most bail bond agencies require the defendant (or family or friends representing the defendant) will also need to pay a significantly larger

deposit or may be required to sign over collateral for the full bail amount (Williams 2017, Hopkins and Doyle 2018). Bail is often set upwards of several thousand dollars, even for misdemeanor crimes, and many defendants are not able to afford the 10% fee or do not own anything worth the needed amount of collateral (Natapoff 2011, Sacks and Ackerman 2012).Wealthy individuals are able to purchase their freedom, often regardless of how high bail is set. For instance, despite the severity of his accused crimes, Harvey Weinstein was able to post a $1 million bond within only hours of his arrest (Kozlowska 2018). This is in stark contrast to more than 500,000 defendants currently being detained because they could not afford their bail amount (Rabuy and Kopf 2016).

Implications of Pretrial Detention

The negative implications of pretrial detention begin immediately and have lasting consequences. First, those who are detained prior to their trial are significantly more likely to be convicted and to serve a longer sentence compared to those who were released prior to their trial. For example, Heaton, Mayson, and Stevenson (2016) found that misdemeanor defendants who were detained were not only 25% more likely to plead guilty but also 43% more likely to receive jail time, which is, on average, twice the length of those who were granted a pretrial release. Being detained decreases the defendant's ability to aid their defense in building a case for their release. For instance, the defendant is only permitted a small amount of time for calls and visits that must be also be used to talk with family. Detained defendants are also unable to visit the scene of the accused crime to help their lawyer recreate the scene or discuss important details (Bender, Lum,

and Wilkerson 2018). Similarly, due to the harsh conditions of imprisonment, as well as the outside consequences of being detained, such as the loss of housing or employment, the defendant may be incentivized to plead guilty to speed up the eventual release. This worsens the already asymmetric power dynamics during plea negotiations, leading detained defendants to plead guilty regardless of guilt (Dobbie, Goldin, and Yang 2018).

A criminal record, even for a misdemeanor, can have life-altering consequences after the defendant's release. These include heavy fines, deportation, loss of housing, loss of employment, and loss of child custody, to name only a few (Heaton, Mayson, and Stevenson 2016). Convictions can also have downstream ramifications, such as limitations on future educational, employment, and social opportunities (Natapoff 2011). This also has the consequence of substantially lowering the defendant's lifetime income and, when considered with the hundreds of thousands of individuals who are or have been imprisoned, also ultimately decreases tax revenue (Dobbie, Goldin, and Yang 2018). In addition to this loss in future revenue, the government also spends an estimated 13.6 billion dollars on pretrial incarceration (Rabuy 2017).

Pretrial Risk Assessments

The costs of pretrial detention to individuals, communities, and the state as well as evidence of unjust and biased judicial decisions, have led to significant criminal justice reform efforts, many of which include the reform or abolition of monetary bail. Generally, pretrial detainment decisions often go beyond simply analyzing and understanding information; many of these decisions require predicting the likelihood of

future events (e.g., whether if released, the defendant will fail to return to court or if they will be a danger to others). Because it cannot be known what will happen, the judge must extrapolate (Singh & Fazel 2010; Mayson 2019). Furthermore, judges typically have incomplete information available, which they may prioritize or deprioritize based on the factors they deem necessary or most important. The decision of whether or not to detain a defendant, release them on bail (and if so, how much), or on their own recognizance is largely decided based on intuition or other subjective, implicit calculations and fallible cognitive processes (DeMichele et al. 2018).

In contrast, decision aiding software has been seen as a way to provide judges with information based strictly on statistical probabilities of risk which have been determined through analysis of prior cases. These tools use a defendant's information, which vary but can include socioeconomic indicators, information about the current case and previous criminal history. These systems calculate a risk score and/or recommendation to help judges assess the likelihood of certain behaviors, usually related to either risks the defendant poses (flight risks, risk of future violence, etc.) or needs they may require (such as addiction or mental health services) (Desmarais and Lowder 2019). The aim of pretrial risk assessment instruments is to influence or "guide" judges to ground release decisions in more objective and predictive factors than either the charge alone or information gleaned from brief interviews or legally fraught or protected characteristics (e.g., race, gender, appearance, mannerisms, etc.). DeMichele et al. (2017) argue that risk assessment systems allow judges to review and override their initial perceptions of a case given further information. This perspective aligns with research by

behavioral economists that assert that individuals can be "nudged" into making better decisions. This "libertarian paternalism" holds that individuals ultimately retain discretion, but institutions can shape environmental factors to influence decisions in a way that systematically guide individuals to make decisions that produce more socially beneficial outcomes without mandating that behavior (Thaler and Sunstein 2009).

Risk assessment instruments take a set of inputs often including demographic factors, current charge(s), and criminal history, employment status, residence, community ties, and drug use (Hannah-Moffat 2015, Kennealy 2018). Some assessment algorithms also include information derived from a personal interview or survey, including questions such as "[h]ow many of your friends/acquaintances are taking drugs illegally?" and "[a] hungry person has a right to steal" (Angwin et al. 2016). These data are then used to produce one or more risk scores, often on a numerical scale (i.e., scores of 1-10) or a classification scale (i.e., scores of "low, moderate, or high" risk). The majority of these tools have proprietary processes. For instance based on Northpointe's public disclosure, we know that "COMPAS is an algorithm that analyzes a number of demographic, social, and physiological factors in a publicly unknown way (its underlying processes and weights are proprietary information) to derive an estimate for how much of a 'risk' an arrestee is for fleeing from their court date or committing a new crime if released on bail" and that "it consists of predictive risk scales for risk prediction" (4-5). Various risk assessments have been adopted on both the federal and state level (Wreth 2019).

Adoption and implementation of pretrial risk assessment tools.

Various risk assessment frameworks have been used in the criminal justice system off and on since the 1920s. These processes and tools were often intended to predict criminality, but the manner in which they do this has varied over the years. The first approaches to assessing the risks posed by a defendant occurred in the 1920s (Harcourt 2008). This was an unstructured approach that encouraging a professional assessment by a clinician to assess criminality using factors like "the look in a prisoner's eye" (Solow-Niederman, Choi, and Van den Broeck 2019; also see Rothman 2017). The subjective nature of this approach drew criticism and, by the 1960s, spurred the creation of more structured tools, such as the Vera Institute's Manhattan Bail Project risk assessment, that produced simplistic scores based on a small list of factors. Various other risk assessments were deployed throughout the years, though few that were empirically validated or used beyond one state or county (Lowenkamp, Lemke, and Latessa 2008). More rigorous statistical and computational risk assessment tools, including those using machine learning techniques, began in the early 2000s in various states and jurisdictions, including Virginia and Washington D.C.. By 2009, the Administrative Office of the U.S. Courts recommended the development of an actuarial pretrial risk assessment for federal courts. The "Pretrial Risk Assessment" (PTRA) was implemented nationally in 2011 (Cadigan and Lowenkamp 2011).

Pretrial risk assessment tools are much more varied on the state level, with some mirroring the PTRA, while other states have purchased algorithmic tools from private or tasked government technology offices with designing new aids. This variation is, in part,

due to vastly less stringent technology procurement regulations on the state level, which gives the state and local jurisdictions much more flexibility to request and procure technology (Brown 2014). There is also the practical need to design new risk assessment in order to appropriately analyze local populations that may have unique differences compared to other localities (Latessa et al. 2010, Werth 2019). Similarly, models may need to be reformulated based on each jurisdictions' laws.

The manner in which PRAs are mandated with legislation also varies. For instance, The New Jersey Criminal Justice Reform Act of 2017 explicitly adopted the Laura and John Arnold Foundation's[41] risk assessment, the Pretrial Safety Assessment (PSA). This has been one of the most widely adopted risk assessment tools, which as of 2017, had been adopted by about 40 jurisdictions nationwide. Some of these jurisdictions have adopted the PSA as part of state-wide implementation, like New Jersey, while some counties have independently chosen to adopt the algorithm (ACLU New Jersey, 2017). In contrast, the California Money Bail Reform Act of 2018 required a varied approach in which California courts could use any "validated risk assessment… from the list of approved pretrial risk assessment tools maintained by the Judicial Council" (CA SB 10, Sec. 4).

[41] The Laura and John Arnold Foundation transitioned from a 501 C3 Foundation to a Limited Liability Corporation (LLC) called "Arnold Ventures" in 2019 (Piper, 2019).

The Decline of Pretrial Risk Assessment Support

The early 2000s saw a generally optimistic narrative about the way technology would change the world with refrains in conferences and classroom about "technology for the social good" (Cohen 2018). However, by the mid 2010s, there was growing concern over the ethical implications of many emerging technologies among journalists, academics, and policymakers. In May 2016, ProPublica, an investigative news organization, broke a story entitled "Machine Bias," which asserted that "[t]here's software used across the country to predict future criminals. And it's biased against blacks" (Angwin et al. 2016). This article found that people of color were twice as likely to be given high-risk assessment scores despite, on average, being significantly less likely to commit future crimes than their white peers. In doing so, ProPublica helped bring these issues to the broader public and launch dozens of studies, reports, and research institutes and initiatives on risk assessment algorithms, particularly those that are used in the criminal justice system. (Courtland 2018; Solow-Niederman, Choi, and Van den Broeck 2019).

Any tool or process added to the current US criminal justice system is inherently ripe for criticism. The criminal justice system has been shaped by racial animosity and disregard, or even malfeasance, toward the poor and working class, and because predictive decision systems rely on historical data to find pertinent trends and patterns that can be used as indicators in new cases, these biases are deeply entrenched in crime data and statistics (Lum et al. 2014, Rabuy and Kopf 2016, Middlemass 2018). As such, critics of the ADSs in the criminal justice system often make a version of the "garbage in,

garbage out" argument, though it tends to be oriented around personal or systematic errors stemming from prejudice or discrimination, or as Mayson (2019) describes it, "Bias In, Bias Out".

Due to the current injustices and inequalities in policing and the legal system, any proposed remedy will still interact with many varied layers of bias. This is not to say that algorithmic risk assessments cannot provide marginal improvement to a highly flawed system. For instance, a risk assessment algorithm implemented in Virginia in 2004, the first to be used for non-violent offenses, was found to dramatically decrease the rise in nonviolent incarceration rates. Within the first decade of use, incarceration rates were only increasing by about 5%, down from the 31% growth prior to the tool's implementation (Virginia Criminal Sentencing Commission 2014). However, even the most accurate tools will be a part of a flawed system, and thus open to criticism regarding their contribution to harmful systemic outcomes. Virginia's incarceration rate is not growing as rapidly, but it is still continuously growing. As Human Rights Watch notes:

> "Profile based risk assessment, even if it could be made less biased, less arbitrary and more transparent, remains a tool to efficiently process cases, which allows the criminal justice system to continue its current pattern of over-criminalization." (2017, p. 4)

Opponents argue that pretrial risk assessments only provide legitimacy to an unjust system and potentially stymie more meaningful reforms (Human Rights Watch 2017). Generally, critiques of algorithmic risk assessments are inextricably tied to broader systemic and institutional problems within the criminal justice system. the system as a whole, and concerns about the risk assessment, whether intrinsic or extrinsic to the

tool, are tangled within broader, systemic concerns. This narrative has caused a backlash against pretrial risk assessments and have even resulted in criminal justice reform bills being challenged or revoked in the face of wide-spread criticism (Solow-Niederman, Choi, and Van den Broeck 2019).

The focus on pretrial risk assessments make related legislation (i.e., legislation that includes an algorithmic risk assessment) an ideal case study to analyze the effects of both positive and negative narratives on bipartisan support and bill passage. However, given the unique lifecycle of this specific kind of decision aiding system, the trends discussed within this chapter are not necessarily representative of trends of other Algorithmic Decision Systems, such as the systems used to determine welfare eligibility and benefits. That said, information gleaned from risk assessments in the criminal justice system may provide insight regarding the effects of a tool becoming more scrutinized or politicized (see Whittaker et al. 2018).

Data and Methodology

The discussion above illustrates both the incentives and political pressure for pretrial risk assessments and the backlash arising from the use and misuse of risk assessment systems within criminal justice. T, so building on the idea that ADSs, like ADSs, may be perceived as being advantageous to policymakers (as illustrated in Chapter 3), this Chapter seeks to 1) explore initial empirical evidence regarding legislative support for ADSs and 2) whether changes in public perception of these algorithms may dampen baseline support. Based on the legislative adoption pattern of

Pretrial Risk Assessments, I develop several tests for these issues below. The next section will discuss the data sources and collection methods used within my analysis, as well as detail my methodological approach. This will be followed by a section that discusses the results from the analysis and their implications. Lastly, I will briefly outline directions for future research.

Data Sources and Collection

The preceding discussion and the discussion in Chapter 2 yield two general types of hypotheses concerning the impact of risk assessments on criminal justice legislation. From Chapter 2, one of the main impacts of algorithmic policy interventions is to create larger and more diverse legislative coalitions in support of bills that contain algorithms. This is due to the uncertainty that surrounds the actual outcomes from the use of algorithms coupled with the 'objectivity' framing that surrounds the use of such tools. Ideologically motivated legislators are likely to believe that the 'objective' truth generated by an algorithmic process will more likely match their preferred outcomes; this in turn helps increase the 'win set' of potential agreements between ideologically diverse veto point holders, i.e., legislators with different outcome preferences will be willing to agree on the same policy with the inclusion of an algorithm under the belief that the algorithm will produce outcomes more in line with their preferences. Risk assessments, as a type of algorithm, should benefit from this baseline benefit in coalition building. This should be observable in data that examines the bipartisan support of legislation, the raw

count of co-authors on a piece of legislation, and eventual enactment of legislation, under the assumption that the prior coalitional benefits aids in the passage of legislation.

However, the prior discussion in this chapter indicates that over time, a substantial backlash has built against the use of risk assessment algorithms in criminal justice legislation. This indicates that the baseline coalitional benefits from risk assessments should erode in more recent years, possibly resulting in overall negative effects from inclusion of a risk assessment algorithm. Testing this notion requires the use of data with some type of time component, to gain traction on how the relationship between algorithms and the dependent variables of interest (bipartisan support, co-author count, and eventual enactment) has changed over time. This results in two general classes of hypotheses for evaluation: one that focuses on the baseline effects of inclusion of an algorithm component and a second that focuses on how the effects of algorithmic policy inclusion changes over time.

Fortunately, the National Conference of State Legislatures has collected records of state-level criminal justice legislation in two separate databases that can be used to evaluate these classes of hypotheses. One covers criminal justice legislation passed in each state legislature for each year between 2012 and 2017. These data allow for the evaluation of both the baseline effects and the time dependent effects of algorithmic policy inclusion, but only on legislation that successfully passed through a state legislature. This has two drawbacks, 1) the impact of algorithmic inclusion on actual passage of legislation cannot be observed with these data, only generally coalitional building effects, and 2) the data only allow for analyzing the impact of algorithmic

inclusion on legislation that was ultimately passed; it does not provide a full accounting of information at the beginning of the introduction state for a piece of legislation. The second database contains information on criminal justice legislation introduced on the state level for 2018. These data allow for the explicit evaluation of the relationship between inclusion of a risk assessment algorithm and eventual policy enactment, in addition to the coalitional effects, as well as providing insight earlier in the policy making process by including legislation that was introduced but ultimately failed. However, these data do not allow for the evaluation of time dependent effects, and given that it only covers the most recent year, i.e., where the backlash effect is likely strongest, the baseline effect for inclusion of a risk assessment algorithm is likely to be at its smallest, or potentially its most negative, value.

Data were collected from the National Conference of State Legislatures' (NCSL) database containing enacted pretrial policy legislation from 2012-2017, and failed and enacted pretrial policy legislation in 2018. The database contains the following information: the policy name and number, the date of last action, a summary of the bill, bill authors and their respective political affiliations; and tags categorizing 12 policy areas relating to pretrial reforms. The policy categories include: Bond Forfeiture and Conditions Violations; Budget, Oversight, and Administration; Citation in Lieu of Arrest; Commercial Bond Regulation; Conditions of Pretrial Release; Court Guidance for Release Determinations; Diversion Programs; Eligibility for Pretrial Release; Pretrial Services and Programs; Risk Assessment; Specialized Populations; and Victim Protections and Policy (NCSL 2018).

The data from the NCSL covers the entirety of all successfully passed pretrial legislation in all U.S. states between 2012 and 2017 and the entirety of all introduced pretrial legislation in all U.S. states in 2018. In this sense, the datasets observed for the analysis are the population level data for the respective policy types. The primary goal of the analysis is to demonstrate the viability of observing algorithmic impacts on the legislative process through the lens of risk assessment tools and pretrial legislation. This is aided by having access to population data because it simplifies the statistics involved with such an analysis. The primary implication regards significance testing. Significance testing quantifies the probability of rejecting a true null hypothesis due to sampling error; however, since the data are population level, there is no sampling involved for sampling error to be the reason for false inferential reasoning. Instead, the parameter estimates obtained by the statistical models provide the true population estimates because the data used is the true population data.

However, many researchers still utilize significance testing even while working with population-level data under the rationale of super-populations. The idea of a super-population is that any finite population of size N is actually a sample drawn from an infinite super population (Hartley & Sielken 1975). In this sense, the passed 2012-2017 legislation and the introduced 2018 legislation are just realized samples from two broader super-populations of passed and introduced legislation. The super-population concept allows scholars to generalize their findings from the finite N population to the infinite super-population. This allows scholars to infer that relationships which likely exist in their tested data also are likely to exist in any other population drawn from the same

super population. This can be realized as a relationship in one policy area existing in another, similar policy area, each of which exist in the same policy area across time or jurisdictions. Scholars will therefore use significance testing on population data to increase the generalizability of any potential finding to contexts of temporal persistence or legislative cross-applicability.

The tradeoff for this generalizability is decreased sensitivity, or the ability for the analysis to correctly identify true relations. By subjecting potential findings to significance tests, scholars run the risk or falsely accepting the null hypothesis on the basis of the relevant p-value failing to achieve the selected significance level for the analysis. In the normal pursuit of strong evidence for causal relationships, the risk of a type II error is acceptable in the face of making spurious causal claims.

However, in the context of this specific inquiry into algorithmic interventions, and more specifically pretrial risk assessment algorithms, there is cause to interrogate the rationale and tradeoff of the super-population logic. In particular, there is reason to suspect that the loss in sensitivity is more costly and the gain in generalizability is smaller in the context of risk assessment algorithms. The goal of the empirics is not to make strong causal claims about the impact of risk assessment algorithms on the legislative process. Rather, the goal is to conduct an exploratory analysis to try to detect a relationship between inclusion of algorithmic policy components and various aspects of the legislative process. The literature on empirical evidence between algorithms and legislation is not particularly well developed, so the main goal is to call scholarly attention to the potential impacts, not to make strong causal claims about the existence of

such impacts. In this sense, the risk of committing a type II error is a worse outcome in an exploratory analysis than in an analysis focused on building strong causal evidence.

In addition to the issue of loss of sensitivity, there are also strong reasons to suspect that there is little to gain from generalizing any findings to a theoretical super-population. Risk assessment algorithms were chosen for analysis given that they have a more significant legislative history than most other forms of algorithmic interventions. However, the increased instances of risk assessments in legislation also make them unique: risk assessment algorithms are more visible in the public and the media and they are subject to increased attention from various lobbying and interest groups. Algorithmic interventions are also a fairly new phenomenon, which means that legislators, lobbyists, and bureaucrats are learning how to write and implement laws that involve them. As a result, the super-population from which the observed population data are pulled from is likely to be highly context-dependent. As legislators learn more about how to effectively utilize algorithms, as public and media attention shifts on the subject, and as lobbying and interest group policy goals change, the super-population will also change. Thus, it is unlikely that the super-population will remain the same in different legislative contexts and locations, or even over the next few years. Only by continuing to examine the data over the coming years and in different contexts can scholars begin to build a notion as to the long term impacts of including algorithms in legislation. My analysis is a first step in that process, but as a first step, the tradeoff between sensitivity and generalizability is not the same here as in many other studies. Instead, this analysis will forgo statistical significance testing, and thus the ability to make generalizable claims, for the increased

ability in detecting potential relationships in the data. Since the goal of the analysis is to demonstrate the viability of algorithmic impacts on policy making, the tradeoff in generalizability for explanatory power seems reasonable.

Descriptive Statistics and Variable Discussion

The goal of the empirical analysis is to determine the potential for the inclusion of algorithmic components in legislation to meaningfully impact key points in the legislative process, specifically the co-authorship stage and bill passage. The discussion in Chapter 2 on the potential impact of including algorithmic components in legislation focused not on enactment but on the potential of easing coalition building by blurring perceptions of a policy's outcome and allowing both sides to believe they are getting what they want. In this sense, the most direct causal impact of including an algorithmic component in a piece of legislation is not enactment of the legislation but improvements in the number and diversity of legislator support.

Prior research indicates that bipartisan support is likely to arise when either partisan differences over policy are small or when the uncertainty associated with the policy area is high (Epstein 1998). In the case of ADSs, as argued in Chapter 2, legislators are likely to believe their preferences will be achieved by an algorithmic component because of the vague narrative that the ADS will produce an "objective" outcome. This could produce the practical implication of there being little difference between partisan preferences, even though this is because each party may be misperceiving the likelihood of a potential outcome of an ADS. In short, each side

believes they will get what they want. ADSs are also often employed in complex policy environments, which may also increase overall uncertainty. Furthermore, while polarization has increased overall in Congress, this pattern is much more likely to occur during the floor vote (Harbridge 2011). Thus, bipartisan co-authorship is a relevant indicator of support that is relatively outside of polarized and partisan influences. And while bipartisan support in the form of co-sponsorship may not necessarily result in enactment, it may contribute to the policy content having later legislative success. For instance, broader support may lead to continued reintroduction of the legislation in later legislative sessions or provide resiliency against future legislative changes. Contributions to success beyond each year's session are beyond the scope of this project but should be considered in future research.

The total number of authors may also be indicative of support, though not necessarily bipartisan support. Sponsorship/co-sponsorship is a rare decision that is totally within the control of the legislator (Schiller 1995). It can also be a method to influence the agenda to include a particular issue or policy. Indeed, the more authors on a bill, the more likely it is to receive serious attention from party leadership, in part, because it signals that there are is widespread interest and support for the legislation and that it could be beneficial for many members (Krutz 2005, Bratton and Rouse 2011). Co-sponsorship can also be a signal to other legislators who share goals or have similar constituencies that a bill is of relative importance in comparison to other policies vying for attention (Kingdon 1989, Zhang et al. 2008), potentially garnering further support later in the process (Kessler and Krehbiel 1996; Krutz 2005), or at least building a

legislator's reputation as a proponent of the issue (Schiller 1995). This serves as another way to observe the extent of support for legislation that includes algorithmic decision systems even when that legislation is not enacted.

Enactment serves as a way to observe the effect of ADSs on policy outcomes relative to these other indicators of support. While bill-level attributes, such as multiple sponsors and bipartisan sponsorship, may contribute to eventual policy enactment, the many contextual variables that change based on the specific time period, legislators involved, and content of the bill complicate the relationship between these variables (Bratton and Rouse 2011). This makes enactment a measure of support beyond and, perhaps, significantly different from the support indicated during sponsorship activities. With this in mind, the following empirical analysis examines three main dependent variables of interest and one main independent variable of interest.

Each of the main variables of interest are represented on a bill-year basis: each bill in a specific year is evaluated for whether it meets the criteria for the relative variable. Bills are considered independent across different years, i.e., they are unrelated across year observations. Each variable was calculated from raw data obtained from scrapping the base HTML code from the NCSL pre-trial legislation database webpage. The python code used to parse and construct the dataset is reproduced in the appendix.

Algorithmic- The main IV of interest concerns the inclusion of an algorithmic component in a piece of legislation. The benefit of restricting the analysis to pretrial policy is two fold: 1) the nature of algorithmic policy tools in this space is well defined and limited to pretrial risk assessment tools, and 2) the completeness of the data collected

in the policy area allows for population level analysis. This variable is calculated based on the NCSL topic categorizations, specifically whether or not a piece of legislation was associated with the 'Risk Assessment' topic. This is a binary coded variable where 1 represents the presence of a risk assessment algorithm tool in a piece of legislation.

Bipartisan- This dependent variable represents whether or not a specific bill has an author and at least one co-author from the two major parties, Democrat and Republican; independent legislators are not considered in this calculation. This is a binary coded variable where 1 represents a bill with bipartisan support among the authors and co-authors.

Total Authors- This dependent variable represents the total number of authors attached to a piece of legislation: author, Republican co-authors, and Democratic co-authors; independent legislators are not considered in the analysis.

Enactment- This dependent variable represents whether or not a specific bill was enacted into law. This includes successful passage through both houses of the state legislature (or one in Nebraska), and a signature from the state executive. This is a binary coded variable where 1 represents successful enactment of a piece of legislation. The nature of the enactment variable changes meaning significantly between the 2012-2017 data and the 2018 data. The bills represented in the 2012-2017 data are only the bills that successfully passed through the legislature in any given year in the dataset, i.e., the subset of 2012 data is comprised only of bills that passed the legislature in 2012, the subset of 2013 data is comprised only of bills that passed the legislature in 2013, etc. Once a bill has passed through the legislature the only way it fails to achieve enactment status is

through an executive veto and a corresponding failure to override. In this sense, the enactment variable for the 2012-2017 data more accurately represents the inverse of a successful governor's veto, of which there were 32 (see Table 1).

Figure 1: Number of Criminal Justice Legislation Containing Risk Assessment Components for Each Year 2012-2018

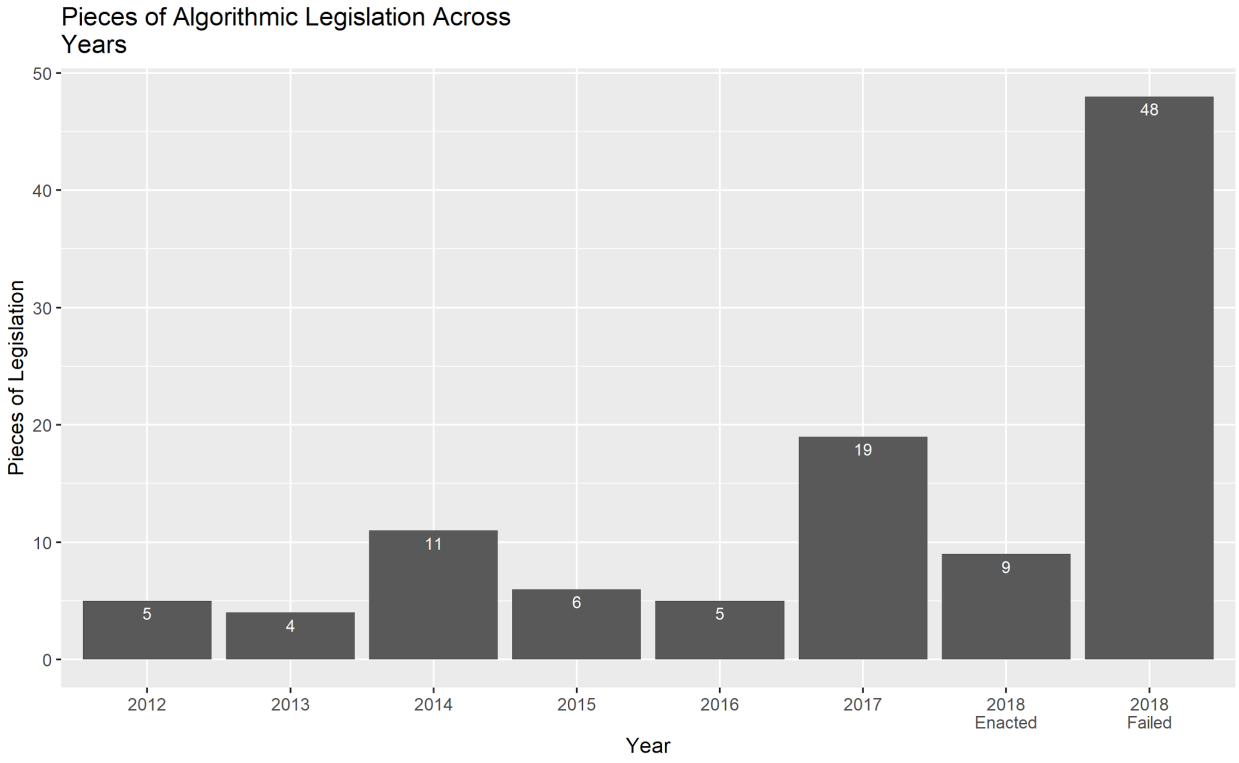

Figure 1 shows the distribution of algorithmic legislation across years in the various datasets. The counts of criminal justice legislation containing risk assessment algorithms are fairly stable for each year of the data set, with the exception of 2017, which saw almost double the number of algorithmic legislation as the next highest count year. The 2018 data is split between bills that were eventually enacted, which are more comparable to the 2012-2017 data, and bills that were introduced but failed at any point

in the process (committee, floor vote, etc.). The only failures included in the 2012-2017 data were bills that were successfully vetoed. There were over five times as many failed pieces of legislation as passed, but it is unknown how this compares to other years in the dataset as there is no information on the number of failed pieces of legislation for the other years in the dataset.

Table 1: 2012-2017 Passed Pretrial Legislation Main Variables Descriptive Statistics

Main Variables of Interest	Algorithmic	Enactment	Bipartisan	Total Authors
N	676	676	676	676
0 Values	626	32	578	0
Min	0	0	0	1
Max	1	1	1	88
Sum	50	644	98	1898
Median	0	1	0	1
Mean	0.07	0.95	0.14	2.81
Var	0.07	0.05	0.12	38.37
Std.Dev	0.26	0.21	0.35	6.19

Table 1 shows descriptive statistics of the main variables of interest for all pieces of enacted pretrial legislation in state governments that were passed during legislative sessions from 2012-2017. Overall, 676 pieces of legislation were passed by state legislatures, with 644, or about 95% being signed into law by a governor. 50 of those bills, or roughly 7%, contained some component of a risk assessment tool. Each piece of legislation had a median of one author indicating that most bill only had a single author.

Only 98 enacted pieces of pretrial legislation over this period, or 14%, had bipartisan support.

Table 2: 2018 Introduced Legislation Main Variables Descriptive Statistics

Main Variables of Interest	Algorithmic	Enactment	Bipartisan	Total Authors
N	641	641	641	641
0 Values	584	495	561	0
Min	0	0	0	1
Max	1	1	1	56
Sum	57	146	80	2157
Median	0	0	0	1
Mean	0.09	0.23	0.12	3.37
Var	0.08	0.18	0.11	38.14
Std.Dev	0.28	0.42	0.33	6.18

Table 2 shows the descriptive statistics for all introduced pretrial legislation in state legislatures in 2018. These data contrast with Table 1 in terms of time (2012-2017 vs. 2018) and scope of included legislation (all passed vs. all introduced). Despite this, many of the findings are similar. The data for introduced legislation in 2018 have slightly higher representation of algorithmic policy tools, in the form of risk assessment algorithms (7% vs. 9%), are slightly less bipartisan (14% vs. 12%), and have on average more total authors associated with each bill (2.89 vs. 3.37), although most pieces of legislation still have a single author with no co-authors. The main difference between the two data sets concerns relative enactment rates. The data collected for 2012-2017 only

reflect bills that had passed, with 95% eventually becoming enacted into law, while 2018 contains the totality of pretrial legislation that was introduced in 2018, of which only 146, or 23%, were enacted into law.

The control variables used for the empirical analysis fall into one of two distinct groupings. Political data and authorship data. Political data refer to the political context a bill was introduced or enacted. Unified control represents whether or not the state government as an entity (legislative and executive branches) was controlled by a single party or was under divided government, coded as 0 or 1, respectively. Democratic and Republican control denote whether either party had complete control over state government, both coded as 0 or 1, respectively. These three variables are closely related to each other. When either Democratic or Republican control is 1, then unified control is also 1. When unified control is 0, so too are Republican and Democratic control. Including all three measures in the empirical analysis allows for the separation of a specific partisan effect from only a unified government effect. More specifically, this allows the model to be sensitive to the difference between legislative styles of Republican majorities and Democratic majorities in the context of criminal justice legislation from the efficiency gains of different veto points being controlled by ideologically similar legislators. There is some concern for multicollinearity in any models given the highly related nature of these variables; however, the highest level of pairwise correlation is only 0.74 (between Unified Control and Republican Control), which is not unduly indicative of a multicollinearity problem. Furthermore, the main impact of multicollinearity is

inefficient parameter estimates, but given that the analyses do not rely on testing for significance, this is less of an issue.

The last political control variable, Party Type, represents the political position of the bill author's party. This variable takes on the values of -1 for a bill authored by a minority party member in the context of unified control, 0 for any bill authored in a divided government context, and 1 for a bill authored by a majority party member in the context of unified control. In this sense, party type is an ordinal variable representing the degree of control of the bill author's party. There are likely other relevant political controls that could be included in the models, but the analysis is restricted to these for two primary reasons. The first is feasibility: additional data on the political context of each state's legislative session for each year would represent a significant resource investment for what would likely be marginal gains in model performance. The second is scope: the goal of the analysis is not to provide rigorous causal evidence for the impact of algorithmic policy tools on the legislative process, but rather to explore the potential for algorithmic impacts and demonstrate potential causal mechanisms and the need for more scholarly attention on collecting and evaluating empirical evidence on the impact of algorithms. In this sense, a well specified model with a complete accounting of relevant controls is less of a necessary requisite for the analysis than being able to provide for an initial account of possible factors, including political influences.

Table 3: 2012-2017 Legislation Political Controls Descriptive Statistics

Political Controls	Unified Control	Republican Control	Democratic Control	Party Type
N	676	676	676	676
0 Values	233	343	566	81
Min	0	0	0	-1
Max	1	1	1	1
Sum	443	333	110	481
Median	1	0	0	1
Mean	0.66	0.49	0.16	0.71
Var	0.23	0.25	0.14	0.37
Std.Dev	0.48	0.5	0.37	0.61

Table 3 displays the distribution of political control attributes for the 2012-2017 passed pretrial legislation. Most bills were passed under the context of unified government (66%) with almost half being passed under unified Republican control (333/676 or 49%), while only 16% 110/676) were passed under unified Democratic control. The median value for party control is 1, indicating that the majority of bills were passed by the majority party. This is true for 71% of pretrial bills passed during this period. This makes the modal bill a Republican authored bill passed during a legislative session marked by unified Republican control.

Table 4: 2018 Legislation Political Controls Descriptive Statistics

Political Controls	Unified Control	Republican Control	Democratic Control	Party Type
N	641	641	641	641
0 Values	257	358	540	86
Min	0	0	0	-1
Max	1	1	1	1
Sum	384	283	101	347
Median	1	0	0	1
Mean	0.6	0.44	0.16	0.54
Var	0.24	0.25	0.13	0.57
Std.Dev	0.49	0.5	0.36	0.76

Table 4 shows descriptive data for all proposed pretrial state legislation in 2018. The numbers here are overall similar to the dataset for passed bills between 2012-2017. Most were passed under unified control, with most of those being characterized by Republican control, although the numbers here are slightly lower (60% vs 66% for unified control and 44% vs 49% for Republican control). The lower average value for Party Type of 0.56 (vs. 0.71 for 2012-2017 data) shows that the bills under consideration are more reflective of minority parties and parties under divided control, which would be expected for a set that includes failed as well as passed bills.

Table 5: 2012-2017 Legislation Authorship Controls Descriptive Statistics

Authorship Controls	Republican Author	Democratic Author	R Co-Author Count	D Co-Author Count
N	676	676	676	676
0 Values	350	439	547	542
Min	0	0	0	0
Max	1	1	48	39
Sum	326	237	554	664
Median	0	0	0	0
Mean	0.48	0.35	0.82	0.98
Var	0.25	0.23	12.63	10.94
Std.Dev	0.5	0.48	3.55	3.31

Table 5 shows the descriptive statistics for authorship variables for passed pretrial state legislation between 2012 and 2017. 83% of passed legislation was authored by either a single Republican or a single Democratic author, with almost half (48%) being written by a Republican author. The median values for both co-author count variables are 0, which reflects the fact that most legislation that was passed had only a single author. That said, Democrats were both more likely to co-author a bill (134 out of 676 cases, or 16.8% of cases had at least one Democratic co-author), and co-authored in larger numbers (a mean value of 0.98 vs. 0.82 for Republicans).

Table 6: 2018 Legislation Authorship Controls Descriptive Statistics

Authorship Controls	Republican Author	Democratic Author	R Co-Author Count	D Co-Author Count
N	641	641	641	641
0 Values	341	396	517	479
Min	0	0	0	0
Max	1	1	52	30
Sum	300	245	650	858
Median	0	0	0	0
Mean	0.47	0.38	1.01	1.34
Var	0.25	0.24	18.58	11.26
Std.Dev	0.5	0.49	4.31	3.36

Table 6 shows the descriptive statistics for introduced pretrial state legislation in 2018. The partisan leanings on authorship are roughly the same as the 2012-2017 data; however, co-authorship was both a more common occurrence, for both parties, as well as occurring in higher counts.

Models and Results

The empirical analysis proceeds in three sections: 1) an analysis of the bipartisan outcome across the 2012-2017 and 2018 data, 2) an analysis of the total number of authors outcome across the 2012-2017 and 2018 data, and 3) an analysis of enactment outcomes across the 2018 data. The 2018 data will be used to explicitly test the direct impact of the algorithm variable on the outcome variables of interest: bipartisan authorship support, total author count, and enactment. The 2012-2017 data will be used to

explicitly test the direct impact of the algorithm variable on the bipartisan and total author count, as well as an interaction term between the year variable and the algorithm variable to test whether the impact of inclusion of algorithmic policy tools within criminal justice legislation has decreased over time. This is done in two separate models: the first, which does not include an interaction between inclusion of an algorithmic component and year, tests for a baseline impact of inclusion of an algorithmic component on the relevant dependent variable while the second, which includes an interaction between inclusion of an algorithmic component and year, tests for a dynamic effect of inclusion of an algorithmic component.

Models 1 and 2: Probability of Bipartisanship

Models 1a, 1b, and 2 explicitly test for the relationship between inclusion of an algorithmic component, operationalized as a risk assessment tool, and the probability of a bill garnering bipartisan support at the authorship stage. Given the nature of the dependent variable as a binary operationalization of whether or not a bill has at least one co-author of the opposite party of the author (or at least one co-author of each party in the case of an independent author), both models rely on logistic regression to obtain parameter estimates. Model 1a and 1b test for a relationship with the 2012-2017 passed legislation while Model 2 tests for a relationship within the 2018 introduced legislation. Inclusion of the 2012-2017 data allows for the additional ability to test if the impact of including an algorithmic component has changed over time, specifically by providing an

additional independent variable constructed by interacting Algorithm with a year variable in Model 1b[42].

The discussion from Chapter 2 on the general nature of algorithmic impacts focused in part on how the uncertainty surrounding specific policy outcomes allows ideologically motivated policy makers to assume their preferred policy outcomes will be met by adopting such 'objectivity-framed' tools. The immediate result of this is an environment more conducive to building bipartisan coalitions, as differing ideologically motivated actors assume that the 'objective' truth will yield a policy outcome closer to their ideal point. However, risk assessment algorithms, as a specific type of algorithmic tool, have had increased scrutiny applied to them over recent years. The combined effect of these two competing influences indicate that including risk assessment algorithms in criminal justice legislation should produce a baseline positive effect that has decreased over time. From this I derive the following two hypotheses concerning the relationship between risk assessment algorithms and criminal justice legislation:

> H1a: Risk assessment inclusion and bipartisan support at bill introduction have a positive association after controlling for authorship and state political control influences
>
> H1b: Interacting risk assessment inclusion and year will produce a negative parameter estimate after controlling for authorship and state political control influences

[42] The year variable here is shifted so that 2012 in the data is coded as 0, so it is better understood as years since 2012.

Table 7: Model 1 and Model 2 Results

	Bipartisan		
	1a	1b	2
Algorithmic	0.92	0.833	0.166
	(0.589)	(1.288)	(0.434)
Republican Author	0.119	0.126	0.504
	(0.263)	(0.263)	(0.249)
Republican Control	-0.559	-0.556	-0.046
	(0.263)	(0.263)	(0.261)
Democratic Control	-0.763	-0.758	-1.202
	(0.411)	(0.411)	(0.451)
Party Type	-0.055	-0.048	-0.341
	(0.208)	(0.209)	(0.158)
Year	0.575	0.593	
	(0.091)	(0.095)	
Algorithmic x Year		-0.228	
		(0.309)	
Constant	-3.439	-3.522	-1.871
	(0.44)	(0.461)	(0.242)
Observations	676	676	641
Log Likelihood	-245.786	-245.537	-232.662
Akaike Inf. Crit.	505.571	507.074	477.323

Table 7 reports the results from Models 1a, 1b, and 2; the primary observation here is that models 1a and 2 both provide support for Hypothesis 1a: inclusion of an algorithmic component is associated with an increased likelihood of bipartisan authorship support in both sets of data. The fact that the data are population level from which no inferences to other populations are being made, means that the standard errors are not of analytical importance for this analysis. Instead, the discussion will be limited to the direction and size of the estimated effects, which, as discussed, are the 'true' population estimates because they are estimated off the true population data.

For the 2012-2017 data, inclusion of a risk assessment algorithm increased the odds of bipartisan support by 151% (exp(0.92) -1) on average over the time period, while the corresponding increase for the 2018 data was 18%. Model 1b provides support for Hypothesis 1b: the relationship between algorithmic inclusion and bipartisan authorship support has decreased with time. Inclusion of a risk assessment tool in a piece of legislation that was eventually passed was associated with an 130% increase in the odds of having bipartisan authorship support, but this was only in 2012. In the following year, 2013, inclusion of a risk assessment tool was associated with only an 83% increase in the odds of having bipartisan authorship support. The impact of risk assessment inclusion would actually switch direction in 2016, where it was associated with an 8% decrease in the odds of having bipartisan authorship support. This relationship obtains its minimal value in 2017, where inclusion of a risk assessment tool was associated with a 26% decrease in the odds of having bipartisan authorship support. The effect size on the algorithmic variable for Model 1b represents the exponentiated value of the coefficient estimated for the algorithm parameter only because in 2012 the year variable takes on a value of 0, so the interaction term between algorithm and year drop out, resulting in the 130% increase in the odds of bipartisan support in 2012 alone. The year variable, when analyzed in isolation, indicates how the odds of bipartisan support changed with each additional year since 2012 when the algorithmic variable is equal to 0, i.e., in bills that did not contain an algorithmic policy tool. The hypothesis is unconcerned with the general trends in bipartisanship for non-algorithmic criminal justice legislation, so the impact of the non-interacted year variable is beyond the scope of this analysis.

The impact of the various control variables was consistent across both models. Both forms of unified control were associated with a decrease in the likelihood of bipartisan support, more so for Democrats than Republicans, relative to a context characterized by divided government. Bills with primary Republican authors were also more likely to attract bipartisan authorship support relative to bills with primary Democratic authors. Bills authored by majority party members under unified government were less likely to attract bipartisan support. For the 2012-2017 data, bipartisan authorship was positively associated with higher years, indicating that bipartisanship increased over time for non-algorithmic pretrial legislation that eventually passed through the legislature.

Models 1 and 2: Results Discussion

The results of Model 1a, Model 1b, and Model 2 suggests there may be aspects of algorithmic decision aids that enable bipartisan coalitions to emerge. However, they also illustrate that, as hypothesized, in this case, coalitional support declined substantially since the analysis began in 2012. This year showed an increase in bipartisan authorship of 85% relative to other pretrial legislation, but by 2017, inclusion of a risk assessment in legislation was associated with a 26% *decrease* in the likelihood of bipartisan authorship. Specifically, the direction of the relationship between risk assessment inclusion and bipartisan support changes in 2016, at the same time criticism of the tools became public and widespread. This indicates that negative public opinion and the rapidly changing

narrative about algorithmic systems in the criminal justice context may depress legislative support, even given the previous structural incentives.

Model 2 results in an increase in bipartisan support in 2018; however, this change in direction, compared to the negative overall relation of risk assessment algorithms in 2017 for Model 1b, is likely due to the fact that this dataset contain all bills (enacted, failed, and pending), so it may be true that the inclusion of risk assessments still increase bipartisan support early in the process, but that many of these bills ultimately fail. While inclusion of an algorithmic component was positively associated with passed legislation for 2012-2015, and then negatively associated in 2016 and 2017, the positive association seen in the 2018 data for Model 2 is not directly comparable since this data includes passed as well as failed pieces of legislation. It may be the case that bipartisan coalitions were still aided by the inclusion of an algorithmic component, but that the broader temporal turn against risk assessments meant that many of these pieces of legislation ultimately failed to pass. Indeed, reanalyzing Model 2 on the subset of 2018 data that did pass shows a negative relationship between inclusion of a risk assessment algorithm and bipartisan support (with a coefficient estimate of -0.59), which continues the time trend observed in the 2012-2017 data.

Models 3 and 4: Total Author Count

Models 3a, 3b, and 4 explicitly test for a relationship between inclusion of an algorithmic policy component in a piece of legislation, operationalized as inclusion of a risk assessment algorithm, and the total number of authors associated with a piece of

legislation. Given the nature of the dependent variable as a count variable, Poisson models are used to estimate model coefficients. Model 3a tests for a relationship with the 2012-2017 passed legislation, Model 3b tests for a dynamic relationship in the 2012-2017 passed legislation, and Model 4 tests for a relationship only within the 2018 introduced legislation. The rationale and expectations are similar to the bipartisan models. The expectation is that there is a baseline positive effect on legislative coalition building, observed as number of co-authors, for including a risk assessment component, but that this effect has eroded over time due to increased scrutiny and public/media skepticism. Therefore, over time, we should observe the inclusion of an algorithmic component be more negatively associated with number of co-authors. This leads to the following hypotheses:

> H2a: Risk assessment inclusion and total author count at bill introduction have a positive association after controlling for author partisanship and state political control influences
>
> H2b: Interacting risk assessment inclusion and year will produce a negative parameter estimate for the impact on total author count after controlling for author partisanship and state political control influences

Table 8: Model 3 and Model 4 Results

	Total Authors		
	3a	3b	4
Algorithmic	-0.399	-0.311	0.087
	(1.04)	(0.272)	(0.073)
Republican Author	-0.181	-0.18	0.044
	(0.051)	(0.052)	(0.047)
Republican Control	-0.111	-0.111	0.101
	(0.052)	(0.052)	(0.051)
Democratic Control	-0.383	-0.383	-0.015
	(0.079)	(0.079)	(0.062)
Party Type	-0.082	-0.081	0.125
	(0.038)	(0.038)	(0.03)
Year	0.237	0.238	
	(0.015)	(0.016)	
Algorithmic x Year		-0.023	
		(0.067)	
Constant	0.529	0.524	1.071
	(0.074)	(0.076)	(0.046)
Observations	676	676	641
Log Likelihood	-2,240.72	-2,240.66	-2,534.20
Akaike Inf. Crit.	4,495.45	4,497.33	5,080.39

Table 8 provides the results for the poisson regressions estimated by models 3a, 3b, and 4. The estimates are reported as coefficient estimates as opposed to incident rate ratios, which aid in the determination of the direction of an effect at the expense of interpretability of effect size. The direction of the effect is the main factor in determining support for a hypothesis for this framework, but making substantive interpretations of the logged rate ratios is difficult. The coefficient estimates (logged rate ratios) can be exponentiated, like log odds in the logit framework, to obtain incident rate ratios that indicate how a one unit change in an independent variable changes the relative ratio of

the count being modeled as the dependent variable. Thus, the estimated coefficient of 0.087 on algorithmic in Model 4 indicates a positive relationship between inclusion of an algorithmic component, operationalized as a risk assessment tool, and the total number of authors associated with a pretrial bill being introduced in 2018. Exponentiating the coefficient yields a value of 1.09, which indicates that a one-unit increase in the algorithmic variable, or the impact of moving from a bill with no algorithmic component to a bill with an algorithmic component, is associated with a 9% increase in the total number of authors[43] associated with a bill. This provides some support for Hypothesis 2a, but the findings from Model 3a provide conflicting evidence. For the 2012-2017 passed legislation, including a risk assessment tool was associated with a 33% decrease (1-exp(-0.399) on average over the time period.

Model 3b provides support for Hypothesis 2b: inclusion of an algorithmic component is associated with a 27% decrease in the total number of authors on that bill (1- exp(-0.311)) in 2012; this decreased over time to a 34% decrease in 2017. Thus, while the overall support for Hypothesis 2a is mixed, Hypothesis 2b does find support with Model 3b; the interaction between algorithm and year is associated with a 1.6% additive decrease per year on the total impact of including an algorithm on total number of authors.

Interestingly, models 3a and 4 also provide conflicting evidence for the impact of the control variables on the total number of authors. The association between unified

[43] In a more technical sense, inclusion of an algorithmic component is associated with a 9% increase in the incident rate ratio, which is the number of relevant analysis units observed over the time period each unit is observed. For the purposes of this analysis, this is the number of co-authors observed over a legislative session, which is one year. This thus simplifies to just the number of authors expected to be observed with a bill.

Democratic control and total number of authors is negative in both the 2012-2017 passed legislation and the 2018 introduced legislation, but the other common controls switch direction in their association on the DV. Being authored by a Republican relative to a Democrat has a negative association with total number of authors in the 2012-2017 data but a positive one in 2018, and the same is true for a bill authored in the context of unified Republican control. Likewise, being authored by a member of the majority party was associated with a decrease in the total number of authors, relative to being authored under the context of divided government, in the 2012-2017 data, but was associated with an increase in the total number of authors in the 2018 data (the inverse would be true for bills authored by minority party members relative to bills authored under the context of divided government). These findings may be indicative of the fact that the data generation process for the total number of authors associated with a bill are fundamentally different for passed legislation between 2012 and 2017 and introduced legislation in 2018 in ways that are not true for bipartisan support at the authorship stage.

Models 3 and 4: Results Discussion

The conflicting results between models 3a and 4 for the impact of risk assessment algorithms provide mixed evidence for algorithmic impacts in legislative policymaking. However, the mirrored results on the control variables between models 3a and 4 indicate the inconclusive evidence for risk assessments is less due to issues with the fundamental theory of algorithmic impacts and more to due to fundamental differences in the data generation processes that underpin the difference between the 2012-2017 dataset and the

2018 dataset. The 2018 dataset represents both passed and unpassed legislation while the 2012-2017 data is comprised only of passed legislation. The 2012-2017 data are characterized by both a lower mean value for total author count and a higher mean value for party type, indicating that the passed legislation for 2012-2017 and the introduced legislation in 2018 differ in that the passed legislation attracted fewer authors and were more likely to be authored by the members of the majority party in the context of unified control. In this situation, the bipartisan coalition building capacity of algorithmic tools could be a liability. Majority party leaders may be less willing to advance bipartisan bills, and majority party members may be less willing to join onto bipartisan legislation. Successful legislation passed by bipartisan coalitions could attract fewer members of the majority party than non-bipartisan coalitions; algorithmic components, by attracting bipartisan support, could decrease the total number of authors by alienating majority party members. Regardless of how the data differ, the almost complete reversal of the signs on the parameters between the two models is less an indication of the mixed results of algorithmic impacts and more about how number of co-authors differs between passed legislation and introduced legislation. These differences are interesting in and of themselves, but beyond the scope of this analysis.

Model 5: 2018 Legislation Probability of Enactment

Model 5 explicitly tests for the relationship between inclusion of an algorithmic component on likelihood of bill enactment by analyzing the relationship between inclusion of a risk assessment algorithm and enactment of pretrial legislation in state

legislatures in 2018. The initial theory discussion would indicate an indirect positive impact of inclusion of a risk assessment algorithm on pretrial legislation enactment, channeled through the positive benefits of building larger coalitions. However, given the significant backlash that risk assessments have received in recent years, coupled with the fact that the data is only for 2018, yields the opposite prediction: that inclusion of a risk assessment algorithm decreases the probability of a pretrial legislation being enacted. The impact of the proposed backlash effect is given increased prominence, effect expectations, and significance specifically due to the fact that the dependent variable is bill passage through the legislature and eventual enactment by the state executive, a particularly strident test of political support. Between the year effects and the particularly high bar for the dependent variable, the baseline positive indirect effects of algorithms on legislative passage discussed in Chapter 2 are not expected to outweigh the effects of popular skepticism and backlash against risk assessments observed in recent years.[44]

[44] There is still, technically, room for a null finding where the baseline effects of algorithmic inclusion washout the backlash effects generated in recent years. There are a number of reasons, specifically in this framework, to suspect this is not the case. The most immediate reason is due to the nature of not engaging in significance testing due to the population level data. A null finding would have to rely on either the effects exactly equaling each other or drawing some arbitrary interval around zero where the combined effects would be considered indistinguishable from zero. In the first case this is just unlikely, and in the second, there is not a pressing reason to draw, and by extension defend, such a threshold when the actual estimates can just be reported and discussed. There are additional, substantive reasons to not suspect a null finding. The discussion in Chapter 2 does not describe a direct link between inclusion of an algorithmic component and bill passage/enactment, rather algorithmic inclusion indirectly impacts bill passage by increasing a legislative coalitions size/diversity and by theoretically increasing the winset under which competing veto players would agree to adopt a new policy over the status quo. This indirect link is already weaker than the direct links expected under bipartisan support and number of co-authors. Additionally, passage and enactment of a bill is overall difficult; only 23% of all introduced legislation was enacted for 2018. Lastly, as the most recent year in the data collected, 2018 is likely to have the strongest effects for the hypothesized backlash, making an already difficult task (enactment) even less likely to occur. Between all these factors, the baseline positive, indirect effect of algorithmic inclusion on legislative passage are very unlikely to exactly match each of the relevant negative factors, some of which are at their hypothesized maximum.

Given the nature of enactment data, this relationship is tested via a logistic regression model with the following hypothesis:

> H3: Risk assessment inclusion and pretrial bill enactment have a negative association after controlling for authorship and state political control influences for the most recent year in the analysis, which follows the peak of negative coverage.

Table 9: Model 5 Results

	Enactment 5
Algorithmic	0.527
	(0.39)
Republican Author	0.292
	(0.226)
Republican Control	0.478
	(0.229)
Democratic Control	1.758
	(0.396)
Party Type	0.631
	(0.175)
Bipartisan	0.663
	(0.271)
Constant	1.369
	(0.214)
Observations	641
Log Likelihood	314.358
Akaike Inf. Crit.	642.716

Table 9 reports the results from the logistic regression of Model 5.The table reports coefficient estimates as log odds to ease in detection of directional impacts. Of immediate note is that the findings on the algorithmic variable are in the negative direction, as predicted. Inclusion of a risk assessment algorithm, was associated with a

decrease in the likelihood of enactment of pretrial legislation by 41% (exp(-0.527)=.59). However, under unified control, Democratic or Republican, the odds of enactment of any pretrial legislation decreases by 83% and 38%, respectively, relative to being introduced in a legislative session characterized by divided control. Meanwhile, authorship by a Republican relative to a Democrat increased the odds of enactment by 34%, bipartisan inclusion of co-authors increased the odds of enactment by 94%, and authorship by a member of the majority party increased the odds of enactment by 88% relative to a member of a party under divided government and 253% relative to authorship by a member of the minority party.

Model 5: Results Discussion

Model 5 shows that there is support for the hypothesis that by 2018, inclusion of an algorithmic component, in the form of a risk assessment algorithm, decreases the likelihood of enactment of a piece of pretrial legislation; in fact, it is likely to decrease the odds of enactment by about 2/5ths. This is in line with the observation within the literature, as reviewed previously, that there was a turn against risk assessment algorithms in the mid/latter part of this decade. It may also be the case that the hypothesized mechanism for which algorithmic policy inclusion impacts enactment, bipartisan coalition building, is a less useful attribute in state legislatures dominated by unified single party control. Enactment of a policy in any given year may have more to do with the demands and goals of party leadership (in the context of single party control of the state government) than with the preferences of bipartisan coalitions. However, the policy

preferences of coalitions are more likely to survive changes in party control and leadership. Analyzing enactment in a single year as a metric for success is unable to capture much of the nuances of the legislative process: proposed legislation can exist in many forms over several legislative sessions. Those bills with bipartisan support may have an advantage in situations of change in party control or movement from unified to divided control. They may even be more robust once enacted, being able to survive changes in party control without significant alteration by a newly elected majority.

As risk assessment algorithms, and algorithms more generally continue to be introduced, marked up, tested against committee votes, amended, tested against floor votes, and sent to governors, better and more plentiful time series data will continue to be generated. As this type of data becomes available, the possibility of testing the relationships between algorithms, bipartisan support, legislative session survival and reintroduction, and eventual enactment becomes more of a reality.

Conclusion

The above analyses indicate that in the early 2010s, there were distinct advantages to including algorithmic components to criminal justice reform legislation, including bipartisan support and attracting greater numbers of authors early in the drafting stage. These advantages may have also given bills containing an algorithmic risk assessment a greater probability of passage compared to other criminal justice reform bills. However, due to the limitations of the NCSL data, this analysis cannot assess implications on bill passage between 2012-2017. The results for 2018, which include all introduced data,

indicated that an algorithmic component in legislation decreases the likelihood of passaged compared to non-algorithmic bills. This is not surprising given the progressive, yearly decline in other kinds of support found in 2012-2017.

One of the main takeaways here is that while there are reasons to suspect, pending more thorough empirical work, that ADSs potentially have systemic advantages, specifically in the form of increased bipartisan support, the potential backlash effects of sustained public scrutiny appear to be enough erode such advantages. This indicates that the potential future of ADSs will be highly dependent on how the public engages with and responds to such systems. To the extent that negative public scrutiny maintains over time, the adoption and spread of ADS systems may be curtailed; however, if these systems improve over time, begin to ameliorate particular harms associated with them, or if public scrutiny recedes, the potential structural advantages of these systems may result in an increase in adoption rates across policy areas.

Fundamentally, this chapter has provided a framework for integrating empirical analyses into broader work on understanding the impact of algorithms on society, while also advancing an argument that algorithmic decision systems are political artifacts that have distinctive political process implications. The experience of pretrial risk assessments provide initial support for the idea that algorithmic policies not only impact but are impacted by distinctly political pressures and need to be studied as explicitly political objects. However, making further claims about the strength of specific causal claims or relative effects of competing impacts will require additional analyses beyond the scope of this project.

Future Research

Future research would benefit from panel data that could analyze all introduced legislation across several legislative sessions to determine whether or not algorithmic policy interventions aided in garnering and maintaining bipartisan support, and whether, in turn, that support aided in a bill surviving through successive legislative sessions until eventual enactment, as opposed to examining enactment at a single point in time. Such data would also allow for post-enactment evaluations of legislative robustness; even after a bill is successfully passed and enacted, it is subject to new legislation curtailing, expanding, or changing its scope. Bipartisan support for enacted legislation may insulate it from changes in partisan composition, priorities, and control in ways more polarized legislation is not. This is just one of many potential avenues of research to pursue; the concluding chapter will, in part, focus on discussing and developing additional methodological pursuits to test and leverage the explanatory power of algorithmic policies in the legislative process.

Chapter 5
Conclusions and Future Research

Algorithmic decision systems have seen a remarkable trajectory in terms of engagement from legislators, bureaucrats, and the public. While the antecedents of these systems have existed for several decades, ADSs in their current form, characterized by machine learning techniques and access to large swaths of training data, have only existed as a legitimate policy tool since about the beginning of the 21st century. Yet despite that short time frame, they have garnered incredible amounts of attention, including from scholars, politicians, government administrators, journalists, and the public. However, the attention has come with rapid adoption in various policy areas, as well as rapid skepticism and backlash against their use in those same policy areas. For scholars engaged in areas ranging from governance to technology to ethics, ADSs represent a rich subject area for gaining new insight and traction on issues and questions endemic to these fields. Scholars in a variety of cross-sectional socio technical disciplines (including political science), are tasked with understanding how computational algorithms can identify and correct issues related to inequality, inaccuracy, or a lack of transparency, despite the fact that the underlying technology is often incredibly difficult to understand or examine due to its proprietary structure and/or blackbox methodological approach. Social scientists have the opportunity to examine how issues of advanced algorithmic systems might impact core socio-political issues like representation, efficacy, injustice, and a litany of other concepts.

Algorithmic policies and decision systems represent such an important research area because they promise so much: ADSs have the potential to drastically increase the capacity of governments to provide services and lower the costs of inefficiencies in governance, or they may potentially entrench historic injustices in ways that make fighting and reversing them once entrenched all the more fraught and difficult. These impacts, and almost every permutation in between, are often the urgent subject of popular and scholarly discourses around the topic of algorithms and society. Whether or not ADSs can deliver on their proponents' loftiest promises, or their detractors' worst fears, remains to be seen, but what is apparent is that governments, local, regional, and national, seem inclined to pass and implement. The rapid change that the implementation of these tools creates is an incredibly important research opportunity, both for bringing new perspectives on old subjects, and for providing the necessary insight and evidence to guide development, design, implementation, and review of these tools. My dissertation has aimed to contribute three meaningful arguments to the scholarly discussion on ADSs.

First, I have sought to elucidate specific ways that ADSs, long considered 'objective' policy tools that simply provide uncontestable truths, are distinctly political artifacts. Specifically, I have emphasized that they (A) are tools for advancing specifically political goals, such as enabling legislators to move policy outcomes closer to their preferences (B) have distinctly political impacts on policy outcomes and processes, and (C) are a locus for political activity, such as increased critical scrutiny and backlash by attentive publics against their adoption and implementation.

Second, I have sought to provide a roadmap for how research into algorithms and society can be paired with empirical work to strengthen arguments and claims. My dissertation does this by providing a macro-level analysis of state legislative activity, which can be used to better analyze systemic questions and build a more robust evidence for evaluating claims and understanding impacts.

Lastly, I have sought to bring focus on how ADSs can have distinctive impacts on political processes, that can in turn result in systemic advantages for these policy tools relative to other policy tools. While the evidence gathered here is not sufficient to make strong causal statements, it is enough to demonstrate a case for the existence of such impacts and advantages and to justify a call for more research and scholarly attention in this direction. Attendant on to this point has been a common theme throughout my dissertation: the tension between potential systemic advantages of ADSs and the moderating effect of public scrutiny and backlash. The evidence presented in Chapter 4 indicates that the effect of the backlash serves to potentially overwhelm any systemic advantages of ADSs for the most recent years of the data. However, whether or not this attention can be sustained, or whether designers of these systems can make changes that meaningfully address concerns remains to be seen. What remains is the necessity for scholars to pay attention and document this tension in order to better understand if and how ADSs may continue to propagate through policy areas and levels of government. Beyond the need to merely pay attention to the tension between systemic advantages and public scrutiny, I conclude below with suggestions for future research.

Future Research

How targeted populations are affected, or how accurately algorithms perform, are questions whose answers will vary across contexts, particularly, as new norms and institutions arise from the interplay of algorithmic policies. But to understand how algorithms will fundamentally interact, shape, and conflict with existing institutions, rules, practices, and distributions of resources and power, scholars must understand how these tools impact processes as well as outcomes. For instance, we must begin to answer the following questions:

- How will algorithmic policies change the behavior of lawmakers, bureaucrats, and lobbyists?
- How will algorithmic policies interact with non-algorithmic policies?
- What issue areas are likely to see more algorithmic policy making activity/less activity?
- How will algorithmic policies change notions of efficacy and representativeness among the broader public?

These questions are ones of process that specifically focus on populations that are not the intended target populations of such policies. Answering these questions will give broader insight into how society and politics might change if more of these policies are proposed and enacted. The goal being to answer a central question: how does a society characterized by algorithms throughout its governing structure differ from one that does not?

This includes the vitally important questions of outcomes: which groups benefit, which are harmed, and how does the distribution of power and resources change? But it also includes equally vital questions of process: how are politics practiced differently, how are resources distributed, how representative and responsive are institutions?

The empirical analysis in the prior chapter provides a roadmap that can be improved upon. One way is through the further collection and analysis on panel data regarding legislation. Bills can linger through multiple sessions of a legislature before being passed or failing to be reintroduced in following session. However, legislation can also garner small and large changes that fundamentally alter the structure and nature of the bill. Pairing panel data on bill introductions and eventual enactments with natural language processing (NLP) techniques can help to identify core policy concepts that make up pieces of legislation and change over time (in this way, we can conceptualize a bill at any given time as a coalition of policy concepts, where the right combination of concepts and context results in the ultimate passage of a piece of legislation). Identifying such core concepts and analyzing how they manifest across legislative sessions can potentially provide significant insight into how certain ideas, tactics, or contexts impact the policy making process. Such a methodological process can also give insight into how pieces of legislation are changed and challenged even after enactment, or why some ideas continue to be reintroduced after failing to pass while others are not. One of the core policy concepts of interest for this research program would be algorithmic policy, in order to gauge how inclusion of such a tool might cause a piece of legislation to be

reintroduced and eventually passed, and by extension what other policy elements are likely to be paired with algorithms and aided by the passage of the combined legislation.

Lastly, future work would benefit from expanding the scope of the political context that algorithmic policies are introduced within. The analysis here focused purely on partisan control; however, there are many more actors in a legislature than just parties. Information on lobbying efforts, media attention, and public opinion are all relevant pieces of information that could impact the success and effects of algorithmic policies. Indeed, much of the discussion in Chapter 4 focused on the potential backlash against risk assessment algorithms; a backlash made possible by sustained public and media attention. If algorithmic policies offer structural advantages to pieces of legislation (such as larger and more diverse coalitions in support), then these advantages are only checked as long the media and concerned publics sustain focus on the issue. However, as public awareness or sentiments change, or as media or lobbying groups shift efforts and attention to other policies/issues, potential systemic advantages can reassert themselves. Expanding the empirical analysis done here to capture variation in other aspects of political context can help elucidate the impact of mitigating factors and how those factors might ebb and flow relative to more robust structural advantages. I hope my dissertation has provided a stepping stone for future analysis on this important topic.

Bibliography

Acemoglu, D., & Robinson, J. A. (2000). Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective. *The Quarterly Journal of Economics*, *115*(4), 1167-1199.

ACLU New Jersey (2017) PRETRIAL JUSTICE REFORM. https://www.aclu-nj.org/theissues/criminaljustice/pretrial-justice-reform

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. new media & society, 20(3), 973-989.

Angwin, J., Larson, J., Mattu s., & Kirchner, L. (2016) Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, May 23rd, 2016

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, *133*(4), 1885-1932.

Bachrach, P. & Baratz, M. (1962). Two Faces of Power, American Political Science Review, 56, 947-52.

Baig, E. (2018). Who's going to review your college applications – a committee or a computer? USA Today, Dec. 2, 2018.

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. Trends in cognitive sciences, 16(12), 593-605.

Barocas, S. & Selbst, A. D. (2016). "Big Data's Disparate Impact." California Law Review, 104.

Bawn, K. (1995). Political control versus expertise: Congressional choices about administrative procedures. American Political Science Review, 89(1), 62-73.

Bechtel, K., Holsinger, A. M., Lowenkamp, C. T., & Warren, M. J. (2017). A meta-analytic review of pretrial research: Risk assessment, bond type, and interventions. American Journal of Criminal Justice, 42(2), 443-467.

Beetham, D. (1996). Bureaucracy. U of Minnesota Press.

Béland, D., Rocco, P., & Waddan, A. (2016). Reassessing policy drift: Social policy change in the United States. Social Policy & Administration, 50(2), 201-218.

Bender, E., Lum, K. & Wilkerson, T. (2018). Understanding the Context and Consequences of Pre-trial Detention. Fairness, Accountability, and Transparency

Bendor, J., & Meirowitz, A. (2004). Spatial models of delegation. *American Political Science Review*, *98*(2), 293-310.

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2018). Early Detection of Students at Risk–Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods.

Besley, T., & McLaren, J. (1993). Taxes and bribery: the role of wage incentives. The economic journal, 103(416), 119-141.

Bimber, B. A. (1996). The politics of expertise in Congress: The rise and fall of the Office of Technology Assessment. SUNY Press.

Birkland, T. A. (2017). Agenda setting in public policy. In *Handbook of public policy analysis* (pp. 89-104). Routledge.

Bratton, K. A., & Rouse, S. M. (2011). Networks in the legislative arena: How group dynamics affect cosponsorship. *Legislative Studies Quarterly*, *36*(3), 423-460.

Brown, M. K. (1981). Working the Street: Police Discretion and the Dilemmas of Reform. New York: Russell Sage Foundation.

Brown, J. (2014) Bringing Innovation to Procurement. GovTech.com retrieved from: http://www.govtech.com/budget-finance/Bringing-Innovation-to-Procurement.html

Buhmann, A., Paßmann, J., & Fieseler, C. (2019). Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse. Journal of Business Ethics, 1-16.

Cadigan, T. P., & Lowenkamp, C. T. (2011). Implementing risk assessment in the federal pretrial services system. Fed. Probation, 75, 30.

Cadigan, T. P., Johnson, J. L., & Lowenkamp, C. T. (2012). The re-validation of the federal pretrial services risk assessment (PTRA). Fed. Probation, 76, 3.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). AI now 2017 report. AI Now Institute at New York University.

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In Conference on Fairness, Accountability and Transparency (pp. 134-148).

Cohen, N. (2018). The know-it-alls: The rise of Silicon Valley as a political powerhouse and social wrecking ball. Oneworld Publications.

Cohen, T. H., & Austin, A. (2018). Examining Federal Pretrial Release Trends over the Last Decade. *Fed. Probation*, *82*, 3.

Cohen, T. H., & Lowenkamp, C. T. (2019). Revalidation of the federal PTRA: Testing the PTRA for predictive biases. Criminal Justice and Behavior, 46(2), 234-260.

Coldewey, D. (2017) New York City moves to establish algorithm-monitoring task force. Tech Crunch

Corbett-Davies, S., Pierson, E., Feller, A., & Goel, S. (2016). "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear". The Washington Post.

Coren, M. (2012) Fighting Violent Gang Crime With Math. Fast Company

Courtland, R. (2018). Bias detectives: the researchers striving to make algorithms fair. Nature, 558(7710), 357-357.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev., 55, 93.

Crawford, K. (2016). Can an algorithm be agonistic? Ten scenes from life in calculated publics. Science, Technology and Human Values, 41(1), 77–92.

Crawford, K. (2017). The Trouble with Bias. Presented at the Neural Information Processing Systems (NeurIPS) Conference

Dahl, R.A. (1957). The Concept of Power, Behavioral Science, 2, 3, 201-215

DeMichele, M., Comfort, M., Misra, S., Barrick, K., & Baumgartner, P. (2018). The intuitive-override model: nudging judges toward pretrial risk assessment instruments.

Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in US correctional settings. Psychological services, 13(3), 206.

Diakopoulos, N. (2013). ‘‘Algorithmic Accountability Reporting: On the Investigation of Black Boxes.’’ A Tow/Knight Brief. New York: Columbia Journalism School, Tow Center for Digital Journalism.

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. American Economic Review, 108(2), 201-40.

Domingos, P. M. (2012). A few useful things to know about machine learning. Commun. acm, 55(10), 78-87.

Doshi-Velez, F., & Kortz, M. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center for Internet & Society.

Dovidio, J. F., Gaertner S.L., & Kawakami, K. (2002). “Implicit and Explicit Prejudice and Interracial Interaction. ” Journal of Personality and Social Psychology 82(1):62-68.

Dunleavy, P., Margetts, H., Tinkler, J., & Bastow, S. (2006). Digital era governance: IT corporations, the state, and e-government. Oxford University Press.

Elish, M. C., & boyd, d. (2018). Situating methods in the magic of Big Data and AI. Communication Monographs, 85(1), 57-80.

Epstein, D., & O'Halloran, S. (1994). Administrative procedures, information, and agency discretion. *American Journal of Political Science*, 697-722.

Epstein, D. (1998). Partisan and bipartisan signaling in congress. Journal of Law, Economics, & Organization, 183-204.

Eslami, M., Aleyasen, A., Moghaddam, R. Z., & Karahalios, K. (2014). Friend grouping algorithms for online social networks: Preference, bias, and implications. In International Conference on Social Informatics (pp. 34-49). Springer.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, 167-181.

Feeley, M. (2017). The process is the punishment. In Crime, Law and Society (pp. 139-188). Routledge.

Feindt, S. (2019). Detained by Data: A Critical Analysis of the Virginia Pretrial Risk Assessment Instrument.

Gailmard, S. (2002). Expertise, subversion, and bureaucratic discretion. *Journal of Law, Economics, and Organization*, *18*(2), 536-555.

Gailmard, S. (2009). Discretion rather than rules: Choice of instruments to control bureaucratic policy making. *Political Analysis*, *17*(1), 25-44.

Gamble, A. E. (2018). A data-based tool to help judges make bail decisions is making us safer. DesMoines Register.

Gehlbach, S. (2013). *Formal Models of Domestic Politics*. Cambridge University Press.

Gelbach, J. & Bushway, S. (2011). Testing for Racial Discrimination in Bail Setting Using Nonparametric Estimation of a Parametric Model. Available at SSRN: https://ssrn.com/abstract=1990324 or http://dx.doi.org/10.2139/ssrn.1990324

Gillespie, T. (2014). The relevance of algorithms. Media technologies: Essays on communication, materiality, and society, 167, 167.

Gillingham, A. (2013). Automated decisions save time. Business Day (South Africa), Feb 27, 2013.

Grothoff, C. & Porup, J. M. (2016). The NSA's SKYNET program may be killing thousands of innocent people. Ars Technica UK.

Guthrie, C., Rachlinski, J., & Wistrich, A. (2007). Blinking on the bench: How judges decide.

Handler, J. F. (1990). Law and the Search for Community. Philadelphia: University of Pennsylvania Press.

Hannah-Moffat, K. (2015). The uncertainties of risk assessment: Partiality, transparency, and just decisions. Federal Sentencing Reporter, 27(4), 244–247.

Harbridge, L. (2011). Congressional Agenda Control and the Decline of Bipartisan Cooperation. Unpublished manuscript, Northwestern University.

Harcourt, B. E. (2008). Against prediction: Profiling, policing, and punishing in an actuarial age. University of Chicago Press.

Hartley, H. O. & Sielken Jr, R. L. (1975). Super-population for finite population sampling. Biometrics, 411-422.

Heaton, P., Mayson, S., & Stevenson, M. (2017). The downstream consequences of misdemeanor pretrial detention. Stan. L. Rev., 69, 711.

Hissong, R. V., & Wheeler, G. (2017). Demographics and Legal Representation in Pretrial Release.

Hopkins, B., Bains, C., & Doyle, C. (2018). Principles of Pretrial Release: Reforming Bail Without Repeating its Harms. Journal of Criminal Law and Criminology, 108(4), 679.

Horn, M. J., & Shepsle, K. A. (1989). Commentary on" Administrative arrangements and the political control of agencies": Administrative process and organizational form as legislative responses to agency costs. *Virginia Law Review*, 499-508.

Huber, J. D., & McCarty, N. (2004). Bureaucratic capacity, delegation, and political reform. *American Political Science Review*, *98*(3), 481-494.

Human Rights Watch (2017). Q & A: Profile-Based Risk Assessment for Pretrial Incarceration, Release Decisions. hrw.org.

Hutchinson, B., & Mitchell, M. (2019). 50 Years of Test (Un) fairness: Lessons for Machine Learning. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 49-58). ACM.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Kehl, D., Guo, P., & Kessler, S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society

Kennealy, P. J. (2018). Are Pretrial Services Officers Reliable in Rating Pretrial Risk Assessment Tools. *Fed. Probation*, *82*, 35.

Kessler, D., & Krehbiel, K. (1996). “Dynamics of Cosponsorship.” American Political Science Review 90: 555–66.

Kingdon, J. W. (1989). Congressmen’s Voting Decisions. 3d ed. Ann Arbor: University of Michigan Press

Kingdon, J. W. (1993). How do issues get on public policy agendas. Sociology and the public agenda, 8, 40.

Kirchner, L. (2015) Machine Bias: What We Know About the Computer Formulas Making Decisions in Your Life ProPublica

Kozlowska, H. (2018a) Prison Inmates will soon be reading ebooks but that's not a good thing. Quartz.

Kozlowska, H. (2018b). Harvey Weinstein's $1 million get-out-of-jail card shows the injustice of bail. Quartz.

Krause, G. A. (2010). Legislative delegation of authority to bureaucratic agencies. In The Oxford handbook of American bureaucracy.

Krutz, G. S. (2005). "Issues and Institutions: 'Winnowing' in the U.S. Congress." American Journal of Political Science 49: 313–26.

Landsbergen D., & Wolken, G. (2001). Realizing the Promise: Government Information Systems and the Fourth Generation of Information Technology. Public Administration Review. 61, 2, 206–220.

Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). Fed. Probation, 74, 16.

Leslie, E., & Pope, N. G. (2017). The unintended impact of pretrial detention on case outcomes: Evidence from New York City arraignments. The Journal of Law and Economics, 60(3), 529-557.

Lipsky, M. (2010). Street-level bureaucracy: Dilemmas of the individual in public service. Russell Sage Foundation.

Locke, J. (2016). Second treatise of government and a letter concerning toleration. Oxford University Press.

Lowenkamp, C. T., Lemke, R., & Latessa, E. (2008). The Development and Validation of a Pretrial Screening Tool. Fed. Probation, 72, 2.

Lukes, S. (1974). Power: A Radical View. London: Macmillan Press.

Lum, K., Swarup, S., Eubank, S., & Hawdon, J. (2014). The contagious nature of imprisonment: an agent-based model to explain racial disparities in incarceration rates. Journal of the Royal Society Interface, 11(98), 20140409.

Maier, E. J. (2017). Schrodinger's Cell: Pretrial Detention, Supervised Release, and Uncertainty. U. Chi. L. Rev., 84, 1425.

Marmot, M. G. (2004). Evidence-based policy or policy-based evidence? BJM.

Marston, G., & Watts, R. (2003). Tampering with the evidence: a critical appraisal of evidence-based policy-making. The drawing board: An Australian review of public affairs, 3(3), 143-163.

Maruna, S., Dabney, D., & Topalli, V. (2012). Putting a price on prisoner release: The history of bail and a possible future of parole. Punishment & Society, 14(3), 315-337.

Maynard-Moody, S., & Portillo, S. (2010). Street-level bureaucracy theory. In The Oxford handbook of American bureaucracy.

Mayson, S. (2019 Forthcoming). Bias in, bias out. Yale Journal of Law, 128.

Middlemass, K. (2017). Convicted and condemned: The politics and policies of prisoner reentry. NYU Press.

McCombs, M., & Shaw, D. L. (2005). The agenda-setting function of the press. *The Press. Oxford, England: Oxford University Press Inc*, 156-168.

McCubbins, M. D., Noll, R. G., & Weingast, B. R. (1987). Administrative procedures as instruments of political control. *Journal of Law, Economics, & Organization*, *3*(2), 243-277.

McCubbins, M. D., Noll, R. G., & Weingast, B. R. (1989). Structure and process, politics and policy: Administrative arrangements and the political control of agencies. Va. L. Rev., 75, 431.

McKinley, J. (2010). New Plan on School Selection, but Still Discontent. New York Times

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Natapoff, A. (2011). Misdemeanors. S. Cal. L. Rev., 85, 1313.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Noll, R. (1983). The political foundations of regulatory policy. Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics, (H. 3), 377-404.

Northpointe (2015). A Practitioner's Guide to COMPAS Core.

O'Donnell. G. (2010). Democracy, Agency, and the State: Theory with Comparative Intent. *Oxford University Press*.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

O'Reilly, T. (2013). ''Open Data and Algorithmic Regulation.'' In Beyond Transparency: Open Data and the Future of Civic Innovation, edited by Brett Goldstein, 289-300. San Francisco, CA: Code for America Press.

Oleson, J. C., Lowenkamp, C. T., Cadigan, T. P., VanNostrand, M., & Wooldredge, J. (2016). The effect of pretrial detention on sentencing in two federal districts. Justice Quarterly, 33(6), 1103-1122.

Pager, D. (2007). Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration. Chicago, IL: University of Chicago Press.

Paine, T. (2003). Common Sense and Other Writings. Modern library.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Perri, S. (2002). 'Can policy making be evidence based?' MCC Building Knowledge for Integrated Care, vol. 10, no. 1, pp. 3–9.

Piper, K. (2019) Why this billion-dollar foundation is becoming a corporation. Vox.

Prottas, J. M. (1979). People-Processing: The Street-Level Bureaucrat in Public Service Bureaucracies. Lexington, MA: Lexington Press.

Rabuy, B. & Kopf, D. (2016). Detaining the Poor: How Money Bail Perpetuates and endless cycle of poverty and jail time. Prisonpolicy.org

Rabuy, B. (2017). Pretrial detention costs $13.6 billion each year, Prison Policy Institute.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). Mitigating Bias in Algorithmic Employment Screening: Evaluating Claims and Practices. arXiv preprint arXiv:1906.09208.

Redford, E. S. (1958). Ideal and Practice in Public Administration. Birmingham: University of Alabama Press.

Rothman, D. J. (2017). Conscience and convenience: The asylum and its alternatives in progressive America. Routledge.

Sacks, M., & Ackerman, A. R. (2012). Pretrial detention and guilty pleas: If they cannot afford bail they must be guilty. Criminal Justice Studies, 25(3), 265-278.

Selbst, A. (2017). A mild defense of our new machine overlords. Vanderbilt Law Review, 70, 87–104.

Schiller, W. (1995). “Senators as Political Entrepreneurs: Using Bill Sponsorship to Shape Legislative Agendas.” *American Journal of Political Science* 39: 186–203.

Shepsle, K. A. (1992). Bureaucratic drift, coalitional drift, and time consistency: A comment on Macey. JL Econ. & Org., 8, 111.

Skolnick, J. (1966). Justice Without Trial: Law Enforcement in Democratic Society. New York: John Wiley & Sons.

Soss, J., Fording, R., & Schram, S. F. (2011). The organization of discipline: From performance management to perversity and punishment. Journal of Public Administration Research and Theory, 21(suppl_2), i203-i232.

Stevenson, M. T. (2017). Assessing risk assessment in action. Minnesota Law Review, 103, 1–65.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. International Journal of Human-Computer Studies, 52(4), 701-717.

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. Criminal Justice and Behavior, 37, 965–988.

Sunstein, C. R. (2014). *Why nudge?: The politics of libertarian paternalism*. Yale University Press.

Tamanaha, B. (2004). On the Rule of Law: History, Politics, Theory. *Cambridge University Press*.

Tan, Y., & Weaver, D. H. (2009). Local media, public opinion, and state legislative policies: Agenda setting at the state level. *The International Journal of Press/Politics*, *14*(4), 454-476.

Tetlock, P. E. (2017). *Expert Political Judgment: How Good Is It? How Can We Know?-New Edition*. Princeton University Press.

Thaler, R. H., & Sunstein, C. R. (2009). Nudge: Improving decisions about health, wealth, and happiness. Penguin.

Tolbert, D., & Solomon, A. (2006). United Nations Reform and Supporting the Rule of Law in Post-Conflict Societies. Harv. Hum Rts. J., 19, 29.

Van Cleve, N. G., & Lara-Millian, A. (2014). Criminal Justice as a Welfare Handout. Presented at the Interplay of Race, Gender, Class, Crime and Justice, University of California, Irvine School of Law

Virginia Criminal Sentencing Commission. (2014). Annual Report.

Volden, C. (2002). A formal model of the politics of delegation in a separation of powers system. *American Journal of Political Science*, 111-133.

Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2019). Algorithmic Bureaucracy: Managing Competence, Complexity, and Problem Solving in the Age of Artificial Intelligence. Complexity, and Problem Solving in the Age of Artificial Intelligence (February 1, 2019).

Weber, M. (1968). Economy and society: an outline of interpretive sociology. Bedminster Press.

Weinberger, D. (2019) *Everyday Chaos: Technology, Complexity, and How We're Thriving in a New World of Possibility.* Harvard Business Review Press.

Werth R. (2019). Risk and punishment: The recent history and uncertain future of actuarial, algorithmic, and "evidence-based" penal techniques. Sociology Compass. https://doi.org/ 10.1111/soc4.12659

Weingast, B. R. (1984). The Congressional-Bureaucratic System: A Principal–Agent Per-spective (with Applications to the SEC). Public Choice.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J. & Schwartz, O. (2018). *AI now report 2018*. AI Now Institute at New York University.

Williams, M. R. (2017). The effect of attorney type on bail decisions. Criminal Justice Policy Review, 28(1), 3-17.

Workman, S. G., Jones, B. D., & Jochim, A. E. (2011). Policymaking, bureaucratic discretion, and overhead democracy. In The Oxford Handbook of American Bureaucracy. Oxford University Press.

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118-136.

Zhang, Y., Friend, A. J., Traud A.L., Porter, M.A., Fowler, J.H., & Mucha, P.J. (2008). "Community Structure in Congressional Cosponsorship Networks." *Physica A: Statistical Mechanics and Its Applications* 387: 1–8.

Zarsky, T. (2016). The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology & Human Values 41(1): 118–132.

Appendix A: Dataset Generation Code

```
#State Pre-Trial Policy Database Constructor

#This script loads the State Pre-Trial Policy website database housed at the
#National Conference of State Legislature website. A version of the database
#site is loaded using a previously constructed raw text file that contains the
#text from a full search of the pretrial database. This data is formatted, cleaned,
#and relevant outcome variables are created.

#Started: 6/24/19

#Import Statements
import requests
import csv
import pandas as pd
from math import sqrt

#Global Definitions
header = ['id', 'year', 'name', 'Status', 'Date of Last Action', 'Author', 'Additional Authors',
'Topics', 'Summary']

#Function Definitions

def loadFile(filename):
    with open(filename, 'r') as file:
        data = file.read()

    return [info.split('\n') for info in data.split('\n\n')]

def loadStateInfo():
    with open('State Party Data.xlsx', 'rb') as file:
        d = {}
        for year in [str(year) for year in range(2012,2019)]:
            reader = pd.read_excel(file, sheet_name = year, index_col=1)
            d[year] = reader.to_dict('index')
    return d

def classifyItem(data):
    if 'Additional Authors' in data:
        data = data.split(':')
        info = data[1].split(' ')
        info_auth = ' '.join(info[0:-2])
```

```
        return(data[0], info_auth, ' '.join(info[-2:]), data[-1])
    elif ':' in data:
        return data.split(':')
    else:
        return '.'

def classify(data):
    info = {'id' : data[0],
            'year' : data[1],
            'name' : data[2]}

    data = data[3:]
    for item in data:
        item = classifyItem(item)
        if item == '.':
            continue
        if len(item) == 4:
            info[item[0]] = item[1]
            info[item[2]] = item[3]
        else:
            info[item[0]] = item[1]

    return (info['id'], info)

def writeLeg(filename, data):
    with open(filename, 'a') as file:
        writer = csv.writer(file, lineterminator = '\n')
        info = [data.get(var, '-') for var in header]
        writer.writerow(info)

def getTopics():
    d ={}
    for topic in topics:
        d[topic] = lambda x, topic = topic: 1 if topic in x else 0
    return d

def statusCheck(status):
    status = status.split('-')
    status = [item.lstrip(' ').strip(' ') for item in status]
    return '-'.join(status[1:]) if status[0] == 'Enacted' else status[1]

def partyType(data):
    if data['Republican Author'] == data['Republican Control']: return 1
```

```
    elif data['Republican Control'] == data['Democratic Control']: return 0
    elif data['Democratic Author'] == data['Democratic Control']: return 1
    else: return -1

#Database Construction
data_pending = loadFile('leg_results_pending.txt')
data_pending = [item[1:] if 'BILL TEXT LOOKUP' in item [0] else item
                for item in data_pending]

data_pending = [classify(item) for item in data_pending]
data_pending = {item[0]:item[1] if item is not None else item for item in data_pending}

data_enacted = loadFile('leg_results_enacted.txt')
data_enacted = [item[1:] if 'BILL TEXT LOOKUP' in item [0] else item
                for item in data_enacted]

data_enacted = [classify(item) for item in data_enacted]
data_enacted = {item[0]:item[1] for item in data_enacted}

with open('pending_leg.csv', 'w') as file:
    writer = csv.writer(file, lineterminator = '\n')
    writer.writerow(header)

with open('enacted_leg.csv', 'w') as file:
    writer = csv.writer(file, lineterminator = '\n')
    writer.writerow(header)

for leg in data_pending.keys():
    writeLeg('pending_leg.csv', data_pending[leg])

for leg in data_enacted.keys():
    writeLeg('enacted_leg.csv', data_enacted[leg])

#Read Data into Pandas Frame
data_pending = pd.read_csv('pending_leg.csv', encoding='unicode_escape')

data_enacted = pd.read_csv('enacted_leg.csv', encoding='unicode_escape')

#Tabulate Topics and Status Types
topics = []
i = 0
for item in set(data_pending['Topics']):
    if 'and' not in item:
```

```
        for topic in item.split(','):
            topic = topic.lstrip(' ').strip(' ')
            if topic not in topics:
                topics.append(topic)

while i < max([string.count('and') for string in set(data_pending['Topics'])]) +1:
    for item in set(data_pending['Topics']):
        if 'and' in item:
            for topic in topics:
                item = item.replace(topic, '')
            if item.count('and') == 1:
                while item[0] in [' ', ','] or item[-1] in [' ', ',']:
                    item = item.lstrip(' ').strip(' ').lstrip(',').strip(',')
                if item not in topics:
                    topics.append(item)
    i += 1

#Break Out Topics
topicFuncs = getTopics()
for topic in topics:
    data_pending[topic] = data_pending['Topics'].apply(topicFuncs[topic])
    data_enacted[topic] = data_enacted['Topics'].apply(topicFuncs[topic])

with open('keyword_file.txt', 'r') as file:
    keyword_list = file.read().split('\n')

summaries_dict = {}
state_info = loadStateInfo()

for data in [data_pending, data_enacted]:
    data['State'] = data['id'].apply(lambda x:x.strip(' ').lstrip(' ').split(' ')[0])
#Create sub-status category
    data['Final Status'] = data['Status'].apply(lambda x: x.split('-')[0].lstrip(' ').strip(' '))
    data['Substatus'] = data['Status'].apply(statusCheck)
#Create additional Author Variables
    data['Additional Authors List'] = data['Additional Authors'].apply(lambda x: x.split(';'))
    data['Republican Author'] = data['Author'].apply(lambda x: 1 if  "(R)" in x else 0)
    data['Democratic Author'] = data['Author'].apply(lambda x: 1 if "(D)" in x else 0)
    data['R Co-Author Count'] = data['Additional Authors'].apply(lambda x:
x.count("(R)"))
    data['D Co-Author Count'] = data['Additional Authors'].apply(lambda x:
x.count("(D)"))
```

```
    data['Total Authors'] = data['Additional Authors List'].apply(lambda x: len(x) + 1 if
x[0] != '-' else 1)
    data['Algorithmic'] = data['Risk Assessments']
#Create State Control Variables
    data['State Control'] = data.apply(lambda x: state_info[str(x['year'])][x['State']]['State
Control'], axis = 1)
    data['Republican Control'] = data['State Control'].apply(lambda x: 1 if x == 'R' else 0)
    data['Democratic Control'] = data['State Control'].apply(lambda x: 1 if x == 'D' else 0)
    data['Unified Control'] = data['State Control'].apply(lambda x: 1 if x=='R' or x=='D'
else 0)
    data['Party'] = data['Republican Author'] + -1*data['Democratic Author']
    data['Party Type'] = data.apply(partyType, axis = 1)
    data['Bipartisan'] = data.apply(lambda row: 1 if (row['Republican Author'] > 0 and
row['D Co-Author Count'] > 0) or
                                    (row['Democratic Author'] > 0 and row['R Co-Author
Count'] > 0) or
                                    (row['R Co-Author Count'] > 0 and row['D Co-Author
Count'] > 0) else 0, axis = 1)
    data['Bipartisan Factor'] = data.apply(lambda row: 2 * (-1/2 + (1/(1 +
abs((row['Republican Author'] + row['R Co-Author Count'])/row['Total Authors'] -
(row['Democratic Author'] + row['D Co-Author Count'])/row['Total Authors'])))), axis
=1)
    data['Weighted Bipartisan Factor'] = data['Total Authors'] * data['Bipartisan
Factor'].apply(sqrt)
    data['Bipartisan Factor'] = data.apply(lambda row: '' if row['Republican Author'] ==
row['Democratic Author'] else row['Bipartisan Factor'], axis = 1)
    data['Weighted Bipartisan Factor'] = data.apply(lambda row: '' if row['Republican
Author'] == row['Democratic Author'] else row['Weighted Bipartisan Factor'], axis = 1)
#Algorithm Detection Loop
    valid_topics = []
    for case in data.index:
        if data.loc[case,'Risk Assessments'] == 1:
            for topic in topics:
                if data.loc[case,topic] == 1 and topic not in valid_topics:
                    valid_topics.append(topic)

    data['Valid'] = 1
    for topic in topics:
        data['Valid'] = data[topic].apply(lambda x, info = topic: 1 if info not in valid_topics
or x == 0 else 0) * data['Valid']
    data['Valid'] = -1 * (data['Valid'] - 1)
    for topic in [topic for topic in topics if topic not in valid_topics]:
        del data[topic]
```

```
#Re-write Data to csv
data_pending.to_csv('data_pending_analysis.csv', index = False)
data_enacted.to_csv('data_enacted_analysis.csv', index = False)
#End Script
```